CONVERGENCE OF THE LARGEST SINGULAR VALUE OF A POLYNOMIAL IN INDEPENDENT WIGNER MATRICES

GREG W. ANDERSON

Abstract. For polynomials in independent Wigner matrices, we prove convergence of the largest singular value to the operator norm of the corresponding polynomial in free semicircular variables, under fourth moment hypotheses. We actually prove a more general result of the form "no eigenvalues outside the support of the limiting eigenvalue distribution." We build on ideas of Haagerup-Schultz-Thorbjørnsen on the one hand and Bai-Silverstein on the other. We refine the linearization trick so as to preserve self-adjointness and we develop a secondary trick bearing on the calculation of correction terms. Instead of Poincaré-type inequalities, we use a variety of matrix identities and L^p estimates. The Schwinger-Dyson equation controls much of the analysis.

Contents

1. Introduction and statement of the main result

1.1. **Background and general remarks.** As part of a larger operator-theoretic investigation, it was shown in [9] (refining earlier work of [10]) that there are for large N almost surely no eigenvalues outside the ϵ-neighborhood of the support of the limiting spectral distribution of a self-adjoint polynomial in independent GUE matrices. (See [1, Chap. 5, Sec. 5] for another account of that result.) It is natural to ask if the same is true for Wigner matrices. We answer that question here in the affirmative. To a large extent this is a matter of learning to get by without the Poincaré inequality. Now the template for results of the form "no

2010 *Mathematics Subject Classification.* 60B20, 15B52.
Key words and phrases. noncommutative polynomials, singular values, support, spectrum, Schwinger-Dyson equation, Wigner matrices.

eigenvalues outside the support..." was established a number of years earlier in the pioneering work of [3], and moreover the authors of that paper got along without the Poincaré inequality quite well—erasure of rows and columns, classical L^p estimates and truncation arguments sufficed. Moreover they got their results under stringent fourth moment hypotheses. In this paper we channel the separately flowing streams of ideas of [3] and [9] into one river, encountering a few perhaps unexpected bends.

Any discussion of largest eigenvalues of Wigner matrices must mention the classical work [4]. In that paper convergence of the largest eigenvalue of a Wigner matrix to the spectrum edge was established under fourth moment hypotheses and it was furthermore established that in a certain sense fourth moments are optimal.

Our main result (see Theorem 1.2.5 immediately below) is both a "polynomialization" of the main result of [4] and a generalization of the random matrix result of [9]. Roughly speaking, to prove our main result, we let the results of [4] do the hard work of attracting the eigenvalues to a compact neighborhood of the spectrum and then we draw them the rest of the way in by using refinements of tools from both [3] and [9], among them matrix identities, L^p estimates for quadratic forms in independent random variables, and the Schwinger-Dyson equation.

Generalizations of the "no eigenvalues outside the support..." result of [9] were quick to appear and continue to do so. In [20], following up on the earlier results of [10], results in the GOE and GSE cases were obtained, and they revealed a key role for "correction terms" of the sort we expend much effort in this paper to control. In [5], a generalization to non-Gaussian distributions satisfying Poincaré-type inequalities was obtained. In [15], a generalization was given involving polynomials in GUE matrices and deterministic matrices with convergent joint law which, in particular, established various rectangular analogues.

All the works following upon [9] including this one build on two extraordinarily powerful ideas from that paper: (i) a counterintuitively "backwards" way of estimating the error of approximate solutions of the Schwinger-Dyson equation and (ii) the famous linearization trick. We refine both ideas in this paper. The refinements are closely intertwined and involve a gadget we call a *SALT block design*.

We have been significantly influenced by the paper [13] which explored geometry and numerical analysis of the Schwinger-Dyson equation, and which could serve uninitiated readers as an introduction to the use of matricial semicircular elements. We were influenced also by [11] and [12] which developed and applied Girko's notion of deterministic equivalent. The notion of deterministic equivalent is in effect exploited here as well, but, more or less following [9], we simply harvest the needed solutions of the Schwinger-Dyson equation from Boltzmann-Fock space fully formed, thus avoiding iterative schemes for producing solutions.

There has been a lot of progress recently on universality in the bulk and at the edge for single Wigner matrices and sample covariance matrices. Edge-universality results in the single matrix case greatly refine and indeed render obsolete results of "no eigenvalues outside the support.." type, albeit usually under more generous moment assumptions. We mention for example [7] which proves convergence of the law of the suitably rescaled smallest eigenvalue of a sample covariance matrix with non-unity aspect ratio to the Tracy-Widom distribution. Of course many other papers could be mentioned—the area is profoundly active at the moment. It seems likely that similar edge-universality results are true in the polynomial case. From

this aspirational point of view our results are crude. But we hope they could serve as a point of departure.

1.2. **The main result.** We now formulate our main result, which at once generalizes the main result of [4] and the random matrix result of [9]. Notation cursorily introduced here is explained in greater detail in §2.1 below.

1.2.1. *Matrices with noncommutative polynomial entries.* Let $\mathbb{C}\langle\mathbf{X}\rangle$ be the noncommutative polynomial algebra generated over $\mathbb{C}$ by a sequence $\mathbf{X} = \{\mathbf{X}_\ell\}_{\ell=1}^\infty$ of independent noncommuting algebraic variables. We equip $\mathbb{C}\langle\mathbf{X}\rangle$ with an involution by declaring all the variables $\mathbf{X}_\ell$ to be self-adjoint. Given a sequence $a = \{a_\ell\}_{\ell=1}^\infty$ in an algebra $\mathcal{A}$ and $f \in \mathrm{Mat}_n(\mathbb{C}\langle\mathbf{X}\rangle)$, we define $f(a) \in \mathrm{Mat}_n(\mathcal{A})$ by evaluating each entry of f at $\mathbf{X}_\ell = a_\ell$ for all ℓ.

1.2.2. *Free semicircular variables.* Let $\Xi = \{\Xi_\ell\}_{\ell=1}^\infty$ be a sequence of free semicircular noncommutative random variables in a faithful C^*-probability space. Given self-adjoint $f \in \mathrm{Mat}_n(\mathbb{C}\langle\mathbf{X}\rangle)$, let the law of the noncommutative random variable $f(\Xi)$ be denoted by μ_f. The latter is a compactly supported probability measure on the real line which depends only on the joint law of Ξ. See §3.3 below for a quick review of C^*-probability spaces and laws of single operators. See §3.4 for the Boltzmann-Fock space construction which yields a sequence Ξ of the required type embedded in an algebraic setup with further structures useful to us. For extensive discussion of noncommutative probability laws, including joint laws, and background on free probability, see [1, Chap. 5].

1.2.3. *Random matrices.* Let

$$\{\{x_\ell(i,j)\}_{1\leq i\leq j<\infty}\}_{\ell=1}^\infty$$

be an array of independent $\mathbb{C}$-valued random variables. We assume the following for all i, j and ℓ:

(1) The law of $x_\ell(i,j)$ depends only on ℓ and $\mathbf{1}_{i<j}$.

(2) $\mathbb{E}|x_\ell(1,1)|^4 < \infty$ and $\mathbb{E}|x_\ell(1,2)|^4 < \infty$.

(3) $\mathbb{E}x_\ell(1,1) = \mathbb{E}x_\ell(1,2) = 0$ and $\mathbb{E}|x_\ell(1,2)|^2 = 1$.

(4) $x_\ell(1,1)$ is real-valued.

(5) The real and imaginary parts of $x_\ell(1,2)$ are independent.

For positive integers N and ℓ we then construct an N-by-N random hermitian matrix X_ℓ^N with entries

$$X_\ell^N(i,j) = \begin{cases} x_\ell(i,j) & \text{if } i<j, \\ x_\ell(i,i) & \text{if } i=j, \\ x_\ell(j,i)^* & \text{if } i>j \end{cases}$$

and for each fixed N, we assemble these matrices into a sequence $X^N = \{X_\ell^N\}_{\ell=1}^\infty$. In turn, given self-adjoint $f \in \mathrm{Mat}_n(\mathbb{C}\langle\mathbf{X}\rangle)$, let ν_f^N be the empirical distribution of eigenvalues of the random hermitian matrix $f\left(\frac{X^N}{\sqrt{N}}\right)$. We use the notation ν_f^N rather than, say, μ_f^N because we are saving the latter for use in our main technical result, namely Theorem 2.4.6 below.

The next result is essentially well-known and provides context for our main result.

Theorem 1.2.4. *For all self-adjoint $f \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X}\rangle)$, the empirical distribution ν_f^N converges weakly to μ_f as $N \to \infty$, almost surely.*

See [21] or [1, Chap. 5] for background, similar results, and many references. See [16] for an interesting recent approach to the proof of a similar result. For the reader's convenience we give in §2.5.2 below a quick derivation of Theorem 1.2.4 from our main technical result.

Now we can state our main result.

Theorem 1.2.5. *For every self-adjoint $f \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X}\rangle)$ and $\epsilon > 0$, $\mathrm{supp}\,\nu_f^N$ is contained in the ϵ-neighborhood of $\mathrm{supp}\,\mu_f$ for $N \gg 0$, almost surely.*

See §2.5.4 below for the derivation of this result from our main technical result.

The next corollary justifies the title of this paper.

Corollary 1.2.6. *For every $f \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X}\rangle)$,*

$$\lim_{N\to\infty} \left[\!\!\left[f\left(\frac{X^N}{\sqrt{N}}\right)\right]\!\!\right] = [\![f(\Xi)]\!] \quad \text{a.s.}$$

Proof. After replacing f by ff^*, we may assume that f is self-adjoint, and furthermore that $f(\frac{X^N}{\sqrt{N}})$ and $f(\Xi)$ are positive. We then need only show that the largest eigenvalue of $f\left(\frac{X^N}{\sqrt{N}}\right)$ converges as $N \to \infty$ to the largest element of the spectrum of $f(\Xi)$, almost surely. In any case, $\mathrm{Spec}(f(\Xi)) = \mathrm{supp}\,\mu_f$ by the very important Proposition 3.4.5 below and thus

$$[\![f(\Xi)]\!] = \sup \mathrm{Spec}(f(\Xi)) = \sup \mathrm{supp}\,\mu_f.$$

Finally, we have

$$[\![f(\Xi)]\!] \le \liminf_{N\to\infty} \left[\!\!\left[f\left(\frac{X^N}{\sqrt{N}}\right)\right]\!\!\right] \le \limsup_{N\to\infty} \left[\!\!\left[f\left(\frac{X^N}{\sqrt{N}}\right)\right]\!\!\right] \le [\![f(\Xi)]\!] \quad \text{a.s.}$$

by Theorem 1.2.4 on the left and Theorem 1.2.5 on the right. □

1.3. **Outline of the paper and plan of proof.**

1.3.1. *Truncation and reduction steps.* In §2, after introducing general notation and terminology, we make the truncation step common to the proofs of Theorems 1.2.4 and 1.2.5. Then we formulate our main technical result, namely Theorem 2.4.6 below, which concerns L^p-norms of "randomized and corrected" Stieltjes transforms, and from it we derive Theorems 1.2.4 and 1.2.5. The proof of the main technical result then takes up the remainder of the paper.

1.3.2. *The self-adjoint linearization trick.* To launch the "blocky" approach taken in the rest of the paper, at the end of §2 we present a very simple self-adjointness-preserving variant of the linearization trick. In Remark 2.6.4 below we explain how the trick gives access to the Stieltjes transforms considered in Theorem 2.4.6.

1.3.3. *General tools.* In §3 we review elementary topics concerning C^*-algebras and in particular we recall the Boltzmann-Fock space construction. We also consider an *ad hoc* version of the notion of Schur complement in a C^*-algebra and use it to solve an abstract version of the Schwinger-Dyson equation. In §4, after introducing tensor products and norming rules, we write down an ensemble of mostly familiar estimates that we will use in place of the Poincaré inequality.

1.3.4. *$\mathcal{S}$-(bi)linear machinery.* In §5 we introduce a collection of algebraic tools needed to take advantage of the fine structure possessed by the random matrices and operators described in Remark 2.6.4. In particular, we introduce the notion of SALT block design to streamline the self-adjoint linearization trick and we develop a secondary trick for making new SALT block designs from old. The secondary trick is a "bootstrapping" technique indispensable to our study of corrections.

1.3.5. *Study of the Schwinger-Dyson equation.* In §6 we recall the Schwinger-Dyson (SD) equation and we construct solutions of it by using the apparatus of §3. Following [9], we use certain of these solutions to represent the Stieltjes transform $S_{\mu_f}(z)$ figuring in Theorems 1.2.4, 1.2.5 and 2.4.6. See Remark 6.2.4 below. We next introduce a secondary version of the SD equation involving notions introduced in §5 and we show how a solution of it can be extracted from the "upper right corner" of a solution of a suitably chosen (larger and more complicated) instance of the SD equation itself. We then construct our candidate for the correction $\mathrm{bias}_f^N(z)$ figuring in Theorem 2.4.6. See Remark 6.5.5 below. In §7, working in a relatively simple geometry, we refine the idea of [9] for controlling errors of approximate solutions to the SD equation. By means of the secondary trick we will be able to use the estimates of §7 not only to study the convergence of empirical distributions of eigenvalues to their limits, but also to study the limiting behavior of corrections.

1.3.6. *Matrix identities and L^p estimates.* In §8 we present a carefully edited catalog of identities satisfied by objects built out of finite-sized chunks of an infinite matrix with entries which are themselves matrices of some fixed finite size. One among these objects via the self-adjoint linearization trick specializes to the randomized Stieltjes transform $S_{\mu_f^N}(\mathbf{z})$ figuring in Theorem 2.4.6. See Remark 8.2.2 below. We note also that identity (140) of §8.5 is the ultimate source of all the correction terms studied here. In §9 we introduce the block Wigner model and work through a long series of L^p estimates culminating in Theorem 9.4.2 below which converts identity (140) to a crucial approximation. We emphasize that all the arguments and calculations presented in §8 and §9 make sense for Wigner matrices when specialized to the case in which the constituent blocks are copies of $\mathbb{C}$. In many cases the calculations so specialized then run along familiar lines. A reader who has already developed some intuition about Wigner matrices should, we hope, be able to build on that base in order to understand our work.

1.3.7. *Concluding arguments.* Finally, in §10 we combine the tools collected above to complete the proof of Theorem 2.4.6 in relatively short order.

2. The truncation step and the main technical result

In §2.1 we introduce general notation in force throughout the paper. In §2.2 we carry out the truncation step for proving Theorems 1.2.4 and 1.2.5. In §2.3 we recall a method for reconstructing a probability measure from its Stieltjes transform. In §2.4 we formulate our main technical result, namely Theorem 2.4.6 below, and we explain how to check its most important hypotheses efficiently using the classical estimate of [8]. In §2.5, we recover both Theorems 1.2.4 and 1.2.5 from Theorem 2.4.6. The proof of the latter result will then take up the rest of the paper. Finally, in §2.6 we introduce a simple self-adjointness-preserving variant of

the famous linearization trick of [9], thereby banishing nonlinear noncommutative polynomials from further consideration in the main body of the paper.

2.1. **Notation and terminology.**

2.1.1. *General notation.* We use $\vee$ and $\wedge$ for maximum and minimum, respectively. Given a complex number $z \in \mathbb{C}$, let $\Re z = \frac{z+z^*}{2}$ and $\Im z = \frac{z-z^*}{2\mathrm{i}}$, and put $\mathfrak{h} = \{z \in \mathbb{C} \mid \Im z > 0\}$, which is the classical upper half-plane. Let $\mathbb{E}$ denote expectation and let $\Pr$ denote probability. (We save the letters E and P for other purposes.) We write $\mathbf{1}_A$ for the indicator of an event A. Let $\operatorname{supp} \nu$ denote the support of a probability measure ν, and similarly, let $\operatorname{supp} \varphi$ denote the support of a function φ. (Recall that supports are always closed sets.) For any $\mathbb{C}$-valued random variable Z and exponent $p \in [1, \infty]$, let $\|Z\|_p$ denote the L^p-norm of Z, i.e., let $\|Z\|_p = (\mathbb{E}|Z|^p)^{1/p}$ for $p \in [1, \infty)$ and otherwise let $\|Z\|_\infty$ denote the essential supremum of $|Z|$. For a matrix A with complex entries, let A^* denote the transpose conjugate, A^{T} the transpose and $[\![A]\!]$ the largest singular value of A. More generally, we use $[\![\cdot]\!]$ to denote the norm on a C^*-algebra. We denote the spectrum of an element x of a C^*-algebra $\mathcal{A}$ by $\operatorname{Spec}(x)$; Proposition 3.2.3 below justifies omission of reference to $\mathcal{A}$ in this notation.

Remark 2.1.2. We use the not-so-standard notation $[\![\cdot]\!]$ for C^*-norms in order not to collide with the notation $\|\cdot\|_p$ for L^p-norms of random variables. We will in fact have to consider expressions of the form $\|[\![A]\!]\|_p$ rather frequently.

2.1.3. *Algebras and matrices.* An *algebra* $\mathcal{A}$ always has $\mathbb{C}$ as scalar field, is associative, and possesses a unit denoted by $1_{\mathcal{A}}$. (Other notation for the unit may also be used, e.g., simply 1.) Let $\operatorname{Mat}_n(\mathcal{A})$ denote the algebra of n-by-n matrices with entries in $\mathcal{A}$. More generally, let $\operatorname{Mat}_{k\times\ell}(\mathcal{A})$ denote the space of k-by-ℓ matrices with entries in $\mathcal{A}$. The (i, j)-entry of a matrix A is invariably denoted $A(i, j)$ (never A_{ij}). Let $\mathcal{A}^\times$ denote the group of invertible elements of an algebra $\mathcal{A}$, put $\mathrm{GL}_n(\mathcal{A}) = \operatorname{Mat}_n(\mathcal{A})^\times$ (GL for *general linear* group) and for $A \in \operatorname{Mat}_n(\mathcal{A})$, let $\operatorname{tr}_{\mathcal{A}} A = \sum_{i=1}^n A(i, i)$. In the special case $\mathcal{A} = \mathbb{C}$ we write $\operatorname{tr} = \operatorname{tr}_{\mathbb{C}}$. Let $\mathbf{I}_n \in \operatorname{Mat}_n(\mathbb{C})$ denote the n-by-n identity matrix and more generally, given an element $a \in \mathcal{A}$, let $\mathbf{I}_n \otimes a \in \operatorname{Mat}_n(\mathcal{A})$ denote the diagonal matrix with entries a on the diagonal. Given a $*$-algebra $\mathcal{A}$, i.e., an algebra endowed with an involution denoted $*$, and an element $a \in \mathcal{A}$, we say that a is *self-adjoint* if $a^* = a$ and we denote the set of such elements by $\mathcal{A}_{\mathrm{sa}}$. Given a matrix $A \in \operatorname{Mat}_{k\times\ell}(\mathcal{A})$ with entries in a $*$-algebra $\mathcal{A}$, we define $A^* \in \operatorname{Mat}_{\ell\times k}(\mathcal{A})$ by $A^*(i, j) = A(j, i)^*$. In particular, by this rule $\operatorname{Mat}_n(\mathcal{A})$ becomes a $*$-algebra whenever $\mathcal{A}$ is.

2.1.4. *The noncommutative polynomial ring* $\mathbb{C}\langle\mathbf{X}\rangle$. Let $\mathbb{C}\langle\mathbf{X}\rangle$ be the noncommutative polynomial ring generated over $\mathbb{C}$ by a sequence $\mathbf{X} = \{\mathbf{X}_\ell\}_{\ell=1}^\infty$ of independent noncommuting variables. By definition the family of all *monomials*

$$\bigcup_{m=0}^{\infty} \{\mathbf{X}_{i_1} \cdots \mathbf{X}_{i_m} \mid i_1, \ldots, i_m = 1, 2, 3, \ldots\}$$

(including the empty monomial, which is identified to $1_{\mathbb{C}\langle\mathbf{X}\rangle}$) forms a Hamel basis for the vector space underlying $\mathbb{C}\langle\mathbf{X}\rangle$. In particular, $\mathbb{C}\langle\mathbf{X}\rangle = \bigcup_{m=1}^\infty \mathbb{C}\langle\mathbf{X}_1, \ldots, \mathbf{X}_m\rangle$. We equip $\mathbb{C}\langle\mathbf{X}\rangle$ with $*$-algebra structure by the rule $\mathbf{X}_\ell^* = \mathbf{X}_\ell$ for all ℓ. Let S^∞ denote the space of sequences in a set S. Given an algebra $\mathcal{A}$, a sequence $a \in \mathcal{A}^\infty$

and matrix $f \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X}\rangle)$, let $f(a) \in \mathrm{Mat}_n(\mathcal{A})$ denote the matrix obtained by evaluating each entry at $\mathbf{X} = a$ (and evaluating $1_{\mathbb{C}\langle \mathbf{X}\rangle}$ to $1_{\mathcal{A}}$). Note that if $\mathcal{A}$ is a $*$-algebra and $a \in \mathcal{A}^\infty_{\mathrm{sa}}$, then $f(a)^* = f^*(a)$, i.e., the evaluation map $f \mapsto f(a)$ is a $*$-algebra homomorphism. If $\mathcal{A} = \mathrm{Mat}_N(\mathbb{C})$, then we view $f(a)$ as an n-by-n array of N-by-N blocks, thus identifying it with an element of $\mathrm{Mat}_{nN}(\mathbb{C})$.

2.1.5. *Empirical distributions of eigenvalues.* Given an N-by-N hermitian matrix A, the *empirical distribution* of its eigenvalues $\lambda_1 \leq \cdots \leq \lambda_N$ is by definition the probability measure $\frac{1}{N}\sum_{i=1}^N \delta_{\lambda_i}$ on the real line.

2.2. **The truncation step.** We begin the proofs of Theorems 1.2.4 and 1.2.5 by proving the following proposition. In doing this we are imitating the initial truncation step taken in [3].

Proposition 2.2.1. *To prove Theorems 1.2.4 and 1.2.5 we may augment assumptions* (1)–(5) *without loss of generality by the following assumptions holding for every index ℓ:*

$$\mathrm{i}^{-\ell} x_\ell(1,2) \textit{ is real-valued.} \tag{6}$$

$$x_\ell(1,1) = 0 \textit{ for odd } \ell. \tag{7}$$

$$\|x_\ell(1,1)\|_\infty < \infty \;\textit{ and }\; \|x_\ell(1,2)\|_\infty < \infty. \tag{8}$$

We collect several tools before completing the proof in §2.2.10 below. The first tool is simply a couple of standard eigenvalue inequalities.

Lemma 2.2.2. *For $A, B \in \mathrm{Mat}_N(\mathbb{C})_{\mathrm{sa}}$ let $\lambda_i(A)$ and $\lambda_i(B)$ denote the i^{th} largest eigenvalue, respectively. Then we have (i) $\bigvee_{i=1}^N |\lambda_i(A) - \lambda_i(B)| \leq [\![A - B]\!]$ and (ii) the corresponding empirical distributions are within distance $[\![A - B]\!]$ as measured in the Lipschitz bounded metric.*

Recall that the distance of probability measures μ and ν on the real line in the *Lipschitz bounded metric* is the supremum of $|\int \varphi\, d\mu - \int \varphi\, d\nu|$ where $\varphi : \mathbb{R} \to \mathbb{R}$ ranges over functions with supremum norm and Lipschitz constant both ≤ 1. Recall also that the Lipschitz-bounded metric is compatible with weak convergence.

Proof. (i) This is well-known. See [14] or [19]. (ii) For any test function $\varphi : \mathbb{R} \to \mathbb{R}$ with sup norm and Lipschitz constant both ≤ 1, since $|\varphi(x) - \varphi(y)| \leq |x - y|$, we have $\left|\int \varphi d\mu_A - \int \varphi d\mu_B\right| \leq [\![A - B]\!]$ by part (i) of the lemma. □

2.2.3. *The Bai-Yin model.* Let

$$\{w(i,j)\}_{1\leq i\leq j<\infty}$$

be an independent family of real random variables such that the law of $w(i,j)$ depends only on $\mathbf{1}_{i<j}$. Assume furthermore that $w(1,1)$ and $w(1,2)$ have finite fourth moments and zero means. Let $\sigma = \|w(1,2)\|_2$. Given a positive integer N, let W^N be the N-by-N random real symmetric matrix with entries

$$W^N(i,j) = \begin{cases} w(i,j) & \text{if } i \leq j, \\ w(j,i) & \text{if } i > j. \end{cases}$$

To have a convenient catchphrase, let us call

$$\{W^N\}_{N=1}^\infty$$

the *Bai-Yin model* for Wigner matrices. We have the following fundamental result.

Theorem 2.2.4 ([4, Thm. C])**.** *In the Bai-Yin model* $\{W^N\}_{N=1}^\infty$*, the largest eigenvalue of* $\frac{W^N}{\sqrt{N}}$ *converges to* 2σ *as* $N \to \infty$*, almost surely.*

Remark 2.2.5. By [4, Thm. A], the fourth moment hypothesis in Theorem 2.2.4 cannot be improved while maintaining strong overall assumptions concerning the form of the joint law of the family $\{W^N\}$ and in particular enforcing the identification of W^N with the upper left N-by-N block of W^{N+1}.

Remark 2.2.6. It is trivial but useful to observe that Theorem 2.2.4 continues to hold in the case $\sigma = 0$, i.e., in the case in which W^N is diagonal. In this case the proof is just an exercise in applying the Borel-Cantelli lemma.

Remark 2.2.7. Only real symmetric matrices were treated in [4] but all the arguments carry over to the hermitian case. In particular, Theorem 2.2.4 continues to hold if we replace $\{W^N\}_{N=1}^\infty$ by the slightly altered family

$$\{\{(\mathrm{i}\mathbf{1}_{i<j} - \mathrm{i}\mathbf{1}_{i>j})W^N(i,j)\}_{i,j=1}^N\}_{N=1}^\infty.$$

Just to have a convenient catchphrase (and to avoid introducing yet more notation) let us call the latter the *twisted* Bai-Yin model.

2.2.8. *C-truncation.* Given a $\mathbb{C}$-valued random variable Z such that $\|Z\|_2 = 1$ and $\mathbb{E}Z = 0$, along with a constant $C > 0$, put

$$\rho_C(Z) = \left\|Z\mathbf{1}_{|Z|\leq C} - \mathbb{E}Z\mathbf{1}_{|Z|\leq C}\right\|_2, \quad \theta_C(Z) = \left\|Z\mathbf{1}_{|Z|>C} - \mathbb{E}Z\mathbf{1}_{|Z|>C}\right\|_2,$$

and if $\rho_C(Z) > 0$ put

$$\mathrm{trunc}_C(Z) = (Z\mathbf{1}_{|Z|\leq C} - \mathbb{E}Z\mathbf{1}_{|Z|\leq C})/\rho_C(Z).$$

Note that

$$\rho_C(Z) \to_{C\to\infty} 1 \text{ and } \theta_C(Z) \to_{C\to\infty} 0 \tag{9}$$

by dominated convergence.

Lemma 2.2.9. *Consider again the Bai-Yin model* $\{W^N\}_{N=1}^\infty$*. But now assume that* $\sigma = 1$*. Let* $C > 0$ *be large enough so that* $\rho_C(w(1,2)) > 0$*. Let* $\widehat{W}^N$ *be the result of applying the truncation procedure* trunc_C *to the off-diagonal entries of* W^N *and putting the diagonal entries of* W^N *to* 0*. (i) We have*

$$\limsup_{N\to\infty} \left[\left[\frac{W^N - \widehat{W}^N}{\sqrt{N}}\right]\right] \leq 2(\theta + 1 - \rho) \;\; a.s. \tag{10}$$

where $\theta = \theta_C(w(1,2))$ *and* $\rho = \rho_C(w(1,2))$*. (ii) The analogous statement holds for the twisted Bai-Yin model.*

Proof. Let D^N be the result of putting the off-diagonal entries of W^N to zero. We have in any case a bound

$$\left[\left[\frac{W^N - \widehat{W}^N}{\sqrt{N}}\right]\right] \leq \left[\left[\frac{D^N}{\sqrt{N}}\right]\right] + \left[\left[\frac{W^N - D^N - \rho\widehat{W}^N}{\sqrt{N}}\right]\right] + \left[\left[\frac{(1-\rho)\widehat{W}^N}{\sqrt{N}}\right]\right] \;\; \text{a.s.}$$

The terms on the right almost surely tend as $N \to \infty$ to 0, 2θ and $2(1-\rho)$, respectively, by Theorem 2.2.4. Thus (i) is proved and (ii) is proved similarly. $\square$

2.2.10. *Proof of Proposition 2.2.1.* The permissibility of assuming (6) and (7) is clear—one has only to break the originally given system of matrices into symmetric and antisymmetric pieces, rescale and relabel. We may assume (6) and (7) henceforth. Now fix $f \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X}\rangle)_{\mathrm{sa}}$ and $\epsilon > 0$ arbitrarily. With a large constant $C_\ell > 0$ depending on ℓ, to be aptly chosen presently, let $\widehat{X}_\ell^N$ be the result of applying the truncation operation trunc_{C_ℓ} to the off-diagonal entries of X_ℓ^N and putting the diagonal entries to 0. Let $\hat{\nu}_f^N$ denote the empirical distribution of eigenvalues of $f\left(\frac{\widehat{X}^N}{\sqrt{N}}\right)$. By Theorem 2.2.4, Remark 2.2.7, Lemma 2.2.9 and (9) we can choose constants C_ℓ large enough, depending on f, so that

$$\limsup_{N\to\infty} \left[\left[f\left(\frac{X^N}{\sqrt{N}}\right) - f\left(\frac{\widehat{X}^N}{\sqrt{N}}\right)\right]\right] < \frac{\epsilon}{2} \ \text{ a.s.}$$

By Lemma 2.2.2(i), almost surely for $N \gg 0$, we have that $\mathrm{supp}\,\nu_f^N$ is contained in the $\frac{\epsilon}{2}$-neighborhood of $\mathrm{supp}\,\hat{\nu}_f^N$, and in turn, by Theorem 1.2.5 applied under the additional assumptions of Proposition 2.2.1, almost surely for $N \gg 0$, we have that $\hat{\nu}_f^N$ is contained in the $\frac{\epsilon}{2}$-neighborhood of $\mathrm{supp}\,\mu_f$. Thus Theorem 1.2.5 in the general case follows from the special case considered in Proposition 2.2.1. A similar argument using Lemma 2.2.2(ii) derives Theorem 1.2.4 in the general case from the special case considered in Proposition 2.2.1. □

2.3. **Stieltjes transforms and reconstruction of probability measures.** We recall an important tool and motivate the introduction of the auxiliary upper-half-plane-valued random variable $\mathbf{z}$.

2.3.1. *Stieltjes transforms.* In general, given a probability measure μ on the real line, recall that the *Stieltjes transform* is defined by the formula

$$S_\mu(z) = \int \frac{\mu(dt)}{t-z} \ \text{ for } z \in \mathbb{C} \setminus \mathrm{supp}\,\mu.$$

We use here the same sign convention as (say) in [3] so that $\Im z > 0 \Rightarrow \Im S_\mu(z) > 0$. Recall also that

$$S_\mu(z^*) \equiv S_\mu(z)^* \ \text{ and } \ |S_\mu(z)\Im z| \le 1. \tag{11}$$

In particular, S_μ is real-valued on $\mathbb{R} \setminus \mathrm{supp}\,\mu$.

2.3.2. *The auxiliary random variable* $\mathbf{z}$. Let m be an even positive integer. Let $\mathbf{z}$ be an $\mathfrak{h}$-valued random variable the law of which is specified by the integration formula

$$\mathbb{E}\varphi(\mathbf{z}) = \int_0^\infty \int_{-\infty}^\infty \varphi(x+\mathrm{i}y) \frac{e^{-(x^2+y^2)/2} y^m}{(m-1)!!\pi}\, dx\, dy.$$

Note that $\mathbf{z} = \mathbf{x} + \mathrm{i}\mathbf{y}$ where $\mathbf{x}$ and $\mathbf{y}$ are independent, $\mathbf{x}$ is standard Gaussian, and $\mathbf{y}$ has density that vanishes to order m at 0. We call m the *strength* of the *repulsion* of $\mathbf{z}$ from the real axis. For simplicity we assume that $\Im\mathbf{z} > 0$ holds without exception. In general we allow m to vary from one appearance of $\mathbf{z}$ to the next. Results below involving $\mathbf{z}$ are often stated with hypotheses to the effect that m be sufficiently large. As we will see the exact distribution of $\mathbf{z}$ is not too important. But it is quite important that $\|1/\Im\mathbf{z}\|_p < \infty$ for $p \in [1, m+1)$. Thus, by choosing the strength of the repulsion of $\mathbf{z}$ from the real axis large enough, the random variable $1/\Im\mathbf{z}$ can be made to possess as many finite moments as we like.

The method we will use for reconstructing probability measures from their Stieltjes transforms is codified by the following lemma in which the auxiliary random variable $\mathbf{z}$ enters as a convenience for bookkeeping.

Lemma 2.3.3. *Let $\varphi : \mathbb{R} \to \mathbb{R}$ be infinitely differentiable and compactly supported. Then there exists a function $\Upsilon : \mathbb{C} \to \mathbb{C}$ depending on the strength of repulsion of $\mathbf{z}$ from the real axis with the following properties:*

(I) *Υ is infinitely differentiable and compactly supported. Furthermore, Υ satisfies $\operatorname{supp} \Upsilon \cap \mathbb{R} = \operatorname{supp} \varphi$ and has the symmetry $\Upsilon(z^*) \equiv \Upsilon(z)^*$.*

(II) *For any open set $D \subset \mathbb{C}$ such that $D^* = D \supset \operatorname{supp} \Upsilon$ and analytic function $b : D \to \mathbb{C}$ such that $b(z^*) \equiv b(z)^*$, we have $\Re \mathbb{E} \Upsilon(\mathbf{z}) b(\mathbf{z}) = 0$.*

(III) *For probability measures μ on $\mathbb{R}$, we have $\Re \mathbb{E} \Upsilon(\mathbf{z}) S_\mu(\mathbf{z}) = \int \varphi \, d\mu$.*

The lemma mildly refines a procedure buried in the proof of [1, Lemma 5.5.5].

Proof. We identify $\mathbb{C}$ with $\mathbb{R}^2$ in the customary way. We switch back and forth between writing $x + \mathrm{i}y$ and (x, y) as it suits us. To begin the construction, let $\theta : \mathbb{R} \to [0, 1]$ be an even infinitely differentiable function supported in the interval $[-1, 1]$ and identically equal to 1 on the subinterval $[-\frac{1}{2}, \frac{1}{2}]$. Let m denote the strength of the repulsion of $\mathbf{z}$ from the real axis. Put

$$\Gamma(x, y) = \frac{1}{2\pi} \theta(y) \sum_{j=0}^{m} \frac{(\mathrm{i}y)^j}{j!} \varphi^{(j)}(x),$$

noting that Γ is supported in $\operatorname{supp} \varphi \times [-1, 1]$. Put $\Gamma'(x, y) = \left(\frac{\partial}{\partial x} + \mathrm{i}\frac{\partial}{\partial y}\right) \Gamma(x, y)$, noting that $\Gamma'(z^*) \equiv \Gamma'(z)^*$. The significance of the differential operator $\frac{\partial}{\partial x} + \mathrm{i}\frac{\partial}{\partial y}$ is that it kills all analytic functions, i.e., it encodes the Cauchy-Riemann equations. The sum defining $\Gamma(x, y)$ is contrived so that

$$\Gamma'(x, y) = \frac{1}{2\pi} \frac{(\mathrm{i}y)^m}{m!} \varphi^{(m+1)}(x) \ \text{ for } (x, y) \in \mathbb{R} \times (-\tfrac{1}{2}, \tfrac{1}{2}).$$

Let $\rho(x, y) = \frac{y^m e^{-(x^2+y^2)/2}}{(m-1)!!\pi}$. Then we have $2\Gamma'(x, y) = \Upsilon(x, y)\rho(x, y)$ for some function Υ satisfying (I). For any Borel measurable function $h : \mathbb{C} \to \mathbb{C}$ satisfying $h(z)^* \equiv h(z^*)$ almost everywhere with respect to Lebesgue measure we have

$$\Re \mathbb{E} \Upsilon(\mathbf{z}) h(\mathbf{z}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Gamma'(x, y) h(x, y) \, dx \, dy \tag{12}$$

provided that the integral on the right is absolutely convergent, as follows directly from the definition of Υ. Furthermore, for any compact set $T \subset \mathbb{R}^2$ with a polygonal boundary and analytic function h defined in a neighborhood of T we have

$$\int_T \Gamma' h \, dx \, dy = -\mathrm{i} \int_{\partial T} \Gamma h (dx + \mathrm{i}dy) \tag{13}$$

by Green's theorem and the fact that h is killed by $\frac{\partial}{\partial x} + \mathrm{i}\frac{\partial}{\partial y}$. To prove (II), take T such that $\operatorname{supp} \Gamma \subset T \setminus \partial T \subset T \subset D$ and take $h = b$. Then formulas (12) and (13) yield the result. To prove (III), assume at first that $\mu = \delta_t$ for some real t and hence $S_\mu(z) = \frac{1}{t-z}$. Take T to be an annulus centered at t and take $h = \frac{1}{t-z}$. In the limit as the inner radius tends to 0 and the outer radius tends to ∞, formulas (12) and (13) yield the result. Finally, to get (III) in general, use Fubini's theorem—the hypotheses of the latter hold by (11) and the fact that $m \geq 1$. $\square$

2.4. **The main technical result.** Now we introduce a new model, the one we actually study through most of the paper, and we formulate a general statement about it, namely Theorem 2.4.6 below. All the hypotheses for this model are ones we deserve to make after performing the truncation step in the proofs of Theorems 1.2.4 and 1.2.5. Furthermore, in certain respects, hypotheses are actually weakened in comparison to those for Theorems 1.2.4 and 1.2.5.

2.4.1. *Data.* For integers $\ell, N \geq 1$, fix a random element Ξ_ℓ^N of $\mathrm{Mat}_N(\mathbb{C})_{\mathrm{sa}}$. Fix also an independent family $\{\mathcal{F}(i,j)\}_{1\leq i\leq j<\infty}$ of σ-fields. Let $\mathcal{F}$ denote the σ-field generated by all the $\mathcal{F}(i,j)$.

2.4.2. *Assumptions.* We assume for each $p \in [1,\infty)$ and index ℓ the following:

$$\sup_N \bigvee_{i,j=1}^{N} \left\|\Xi_\ell^N(i,j)\right\|_p < \infty. \tag{14}$$

$$\sup_N \left\| \left[\left[\frac{\Xi_\ell^N}{\sqrt{N}} \right]\right] \right\|_p < \infty. \tag{15}$$

Furthermore, we assume for all indices ℓ and N the following:

(16) Ξ_ℓ^N is the upper left N-by-N block of Ξ_ℓ^{N+1}.

(17) $(\Xi_\ell^N)^{\mathrm{T}} = (-1)^\ell \Xi_\ell^N$.

(18) $\Xi_\ell^N(i,j)$ is $\mathcal{F}(i\wedge j, i\vee j)$-measurable and $\mathbb{E}\Xi_\ell^N(i,j) = 0$ for $i,j = 1,\ldots,N$.

(19) $\left\|\Xi_\ell^N(i,j)\right\|_2 = 1$ for $1 \leq i < j \leq N$.

Finally, we assume that

$$\mathbb{E}\Xi_\ell^N(i,j)\Xi_m^N(i,j) = 0 \quad \text{for } 1 \leq i < j \leq N \text{ and } 1 \leq \ell < m < \infty \tag{20}$$

for all positive integers i, j, N, ℓ and m subject to the indicated constraints.

Remark 2.4.3. While moment assumptions here are extremely generous in comparison to those of Theorem 1.2.5, in certain other respects we have significantly weakened assumptions. Firstly, we do not require the entries $\Xi_\ell^N(i,j)$ to have law depending only on ℓ and the sign of $i-j$. Secondly, we assume somewhat less than strict independence of the matrices Ξ_ℓ^N for fixed N and varying ℓ.

2.4.4. *Random matrices and empirical distributions of eigenvalues.* For each fixed N we form a sequence $\Xi^N = \{\Xi_\ell^N\}_{\ell=1}^\infty \in \mathrm{Mat}_N(\mathbb{C})_{\mathrm{sa}}^\infty$ of random hermitian matrices. Given any self-adjoint $f \in \mathrm{Mat}_n(\mathbb{C}\langle\mathbf{X}\rangle)_{\mathrm{sa}}$, let μ_f^N denote the empirical distribution of eigenvalues of the random hermitian matrix $f\left(\frac{\Xi^N}{\sqrt{N}}\right) \in \mathrm{Mat}_{nN}(\mathbb{C})_{\mathrm{sa}}$.

2.4.5. *The auxiliary random variable* $\mathbf{z}$. We adjoin the auxiliary random variable $\mathbf{z}$ figuring in Lemma 2.3.3 to our model. We assume that $\mathbf{z}$ is independent of $\mathcal{F}$.

Theorem 2.4.6. *Fix $f \in \mathrm{Mat}_n(\mathbb{C}\langle\mathbf{X}\rangle)_{\mathrm{sa}}$ arbitrarily and let μ_f be defined as in Theorem 1.2.5. Then there exists a sequence*

$$\{\mathrm{bias}_f^N : \mathbb{C}\setminus \mathrm{supp}\,\mu_f \to \mathbb{C}\}_{N=1}^\infty$$

of deterministic analytic functions with the symmetry

$$\mathrm{bias}_f^N(z^*) \equiv \mathrm{bias}_f^N(z)^* \tag{21}$$

such that for every $p \in [1,\infty)$ we have

$$\sup_N N^{1/2}\left\|S_{\mu_f^N}(\mathbf{z}) - S_{\mu_f}(\mathbf{z})\right\|_p < \infty, \tag{22}$$

$$\sup_N N^{3/2}\left\|S_{\mu_f^{N+1}}(\mathbf{z}) - S_{\mu_f^N}(\mathbf{z})\right\|_p < \infty, \tag{23}$$

$$\sup_N \left\|\mathrm{bias}_f^N(\mathbf{z})\right\|_p < \infty \;\; \textit{and} \tag{24}$$

$$\sup_N N^2\left\|\mathbb{E}\left(S_{\mu_f^N}(\mathbf{z})\middle|\,\mathbf{z}\right) - S_{\mu_f}(\mathbf{z}) - \frac{\mathrm{bias}_f^N(\mathbf{z})}{N}\right\|_p < \infty, \tag{25}$$

provided the strength of the repulsion of $\mathbf{z}$ *from the real axis is sufficiently great, depending on* p*.*

Once we have deduced Theorems 1.2.4 and 1.2.5 from Theorem 2.4.6 in §2.5 below, the proof of Theorem 2.4.6 will take up the rest of the paper.

Remark 2.4.7. The theorem if true for some sequence Ξ of free semicircular noncommutative random variables in a faithful C^*-probability space is true for all; only the joint law of Ξ is important. Taking advantage of this freedom, we will make a special choice of Ξ below in §3.4 which is adapted to the symmetry present in our model by virtue of assumption (17).

Remark 2.4.8. Fix a point $z_0 \in \mathfrak{h}$ arbitrarily. For any analytic function $g : \mathfrak{h} \to \mathbb{C}$, we can recover the value $g(z_0)$ as the average of $g(z)$ over the disc $|z - z_0| \leq \frac{1}{2}\Im z_0$. Thus statement (22) for, say, $p = 4$ implies that $S_{\mu_f^N}(z_0) \to_{N\to\infty} S_{\mu_f}(z_0)$, almost surely, by Jensen's inequality in conditional form and the Borel-Cantelli lemma. In short, the averaged result stated in Theorem 2.4.6 easily yields pointwise results.

Remark 2.4.9. An explicit if rather complicated description of $\mathrm{bias}_f^N(z)$ in operator-theoretic terms will be developed below. See Remark 6.5.5 below and the discussion of the Schwinger-Dyson equation which immediately precedes that remark. The role played by $\mathrm{bias}_f^N(z)$ below, justified by Lemma 2.3.3 and relation (25), is to make possible an estimate of $\int \varphi\, d(\mu_f^N - \mu_f)$ accurate to an order in $1/N$ sufficiently high so that we can achieve sensitivity to the movement of individual eigenvalues.

Remark 2.4.10. Suppose that in the setup for Theorem 2.4.6 we discard assumptions (14) and (15) and make in their place for every index ℓ the assumption

$$\sup_{N=1}^{\infty} \bigvee_{i,j=1}^{N} \left\|\Xi_\ell^N(i,j)\right\|_\infty < \infty. \tag{26}$$

Then of course assumption (14) holds trivially, but furthermore and importantly, assumption (15) and the bound

$$\left\|\limsup_{N\to\infty}\left[\left[\frac{\Xi_\ell^N}{\sqrt{N}}\right]\right]\right\|_\infty < \infty \tag{27}$$

hold by the classical result of [8] recalled immediately below in a convenient form.

Proposition 2.4.11. *For each $N \geq 1$, let Y^N be a random N-by-N hermitian matrix whose entries on or above the diagonal are independent. Assume furthermore that the entries of the matrices Y^N are essentially bounded uniformly in N and have*

mean zero. Fix any sequence $\{k_N\}_{N=1}^\infty$ of positive integers such that $\frac{k_N}{\log N}\to\infty$ but $\frac{k_N}{N^{1/6}}\to 0$. Then $\sum_N \mathbb{E}\left[\left[\frac{Y^N}{c\sqrt{N}}\right]\right]^{2k_N}<\infty$ for some (finite) constant $c>0$.

Here and elsewhere throughout the paper constants in estimates are denoted mostly by c, C or K. The numerical values of these constants may of course vary from context to context and even from line to line.

Proof. We will use the result of Füredi-Komlós as cast in the form of the combinatorial estimate [1, Lemma 2.1.23]. By the cited lemma, for any constants

$$\frac{c}{2}>K\geq\sup_N\bigvee_{i,j=1}^N\left\|Y^N(i,j)\right\|_\infty,$$

we deduce via "opening of the brackets" and counting of nonzero terms that

$$\begin{aligned}\mathbb{E}\left[\left[\frac{Y^N}{c\sqrt{N}}\right]\right]^{2k_N} &\leq \mathbb{E}\,\mathrm{tr}\left(\frac{Y^N}{c\sqrt{N}}\right)^{2k_N}\\ \leq\ \frac{1}{c^{2k_N}}&\sum_{t=1}^{k_N+1}2^{2k_N}(2k_N)^{3(2k_N-2t+2)}N^t\frac{K^{2k_N}}{N^{k_N}}\\ =\ N&\left(\frac{2K}{c}\right)^{2k_N}\sum_{t=1}^{k_N+1}\left(\frac{2k_N}{N^{1/6}}\right)^{6(k_N-t+1)},\end{aligned}$$

whence the result, since the last expression summed on N is finite. □

Remark 2.4.12. The argument presented immediately after [1, Lemma 2.1.23] gives the analogous result for Wigner-like random matrices whose entries have L^p-norms uniformly under a bound polynomial in p. We do not need the stronger result here for any of our proofs, but we mention it because it easily produces many natural examples of data satisfying the assumptions of Theorem 2.4.6 and furthermore satisfying (27).

2.5. **Deduction of Theorems 1.2.4 and 1.2.5 from Theorem 2.4.6.**

2.5.1. *Common setup for the proofs.* In order to deduce Theorems 1.2.4 and 1.2.5 from Theorem 2.4.6, we may and we do make the additional assumptions stated in Proposition 2.2.1. In turn, in order to apply Theorem 2.4.6, we now take

$$\Xi_\ell^N=X_\ell^N$$

for all N and ℓ and we put

$$\mathcal{F}(i,j)=\sigma\left(\{x_\ell(i,j)\}_{\ell=1}^\infty\right).$$

Then—hypothesis (15) excepted—the data

$$\{\Xi_\ell^N\}\cup\{\mathcal{F}(i,j)\}$$

trivially satisfy all hypotheses of Theorem 2.4.6, e.g., the algebraic assumptions (6) and (7) imply the (anti)symmetry (17). And hypothesis (15) is fulfilled—not so trivially—by Remark 2.4.10. So Theorem 2.4.6 is indeed applicable in the present case. Note that (27) holds as well, either by yet another application of Theorem 2.2.4 or by Remark 2.4.10.

2.5.2. *Proof of Theorem 1.2.4 with Theorem 2.4.6 granted.* By (27), there exists a constant $A > 0$ such that $\operatorname{supp} \mu_f^N \subset [-A, A]$ for $N \gg 0$, almost surely, so we have tightness. By Remark 2.4.8, we have $S_{\mu_f^N}(\mathrm{i} + 1/k) \to_{N\to\infty} S_{\mu_f}(\mathrm{i} + 1/k)$, almost surely, for every integer $k > 0$. The latter statement by standard subsequencing arguments (which we omit) implies that μ_f^N indeed converges weakly to μ_f, almost surely. □

To derive Theorem 1.2.5 from Theorem 2.4.6 we need a final lemma variants of which have long been in use.

Lemma 2.5.3. *Let $\{Y_N\}_{N=1}^\infty$ be a sequence of nonnegative random variables. Assume that*

$$\sup_N N\mathbb{E}Y_N < \infty \quad \textit{and} \quad \sup_N N^{1/2}\|Y_{N+1} - Y_N\|_4 < \infty.$$

Then $Y_N \to_{N\to\infty} 0$, almost surely.

Proof. We have $Y_{\lfloor k^{5/4}\rfloor} \to_{k\to\infty} 0$, almost surely, by the Chebychev inequality and the Borel-Cantelli lemma. Here $\lfloor x\rfloor$ denotes the greatest integer not exceeding x. Put $[N] = \vee_{k=1}^\infty \lfloor k^{5/4}\rfloor \mathbf{1}_{k^{5/4}<N}$. Clearly, we have $Y_{[N]} \to_{N\to\infty} 0$, almost surely. Since $N - [N] = O(N^{1/5})$, we have $\left\|Y_N - Y_{[N]}\right\|_4 = O(N^{-3/10})$ by the Minkowski inequality. Thus $Y_N - Y_{[N]} \to_{N\to\infty} 0$, almost surely, by the Chebychev inequality and the Borel-Cantelli lemma. The result follows. □

2.5.4. *Proof of Theorem 1.2.5 with Theorem 2.4.6 granted.* Take $\mathbf{z}$ to have a strength of repulsion from the real axis large enough so that all statements of Theorem 2.4.6 hold for the given matrix $f \in \mathrm{Mat}_n(\mathbb{C}\langle\mathbf{X}\rangle)_{\mathrm{sa}}$ in the case $p = 4$. As in the proof of Theorem 1.2.4, fix $A > 0$ such that $\operatorname{supp} \mu_f^N \subset [-A, A]$ for $N \gg 0$, almost surely. Fix $\epsilon > 0$ arbitrarily. Fix an infinitely differentiable function $\varphi : \mathbb{R} \to [0, 1]$ with the following support properties:

- φ is identically equal to 1 on $[-A, A]$ minus the ϵ-neighborhood of $\operatorname{supp} \mu_f$.
- φ is supported in some compact set disjoint from $\operatorname{supp} \mu_f$.

For $N > 0$ consider the nonnegative random variable $Y_N = nN \int \varphi d\mu_f^N$ the value of which for $N \gg 0$ bounds the number of eigenvalues of the random hermitian matrix $f\left(\frac{\Xi^N}{\sqrt{N}}\right)$ straying outside the ϵ-neighborhood of $\operatorname{supp} \mu_f$, almost surely. It will be enough to show that $Y_N \to_{N\to\infty} 0$, almost surely. Now by Lemma 2.3.3 and Fubini's theorem, for some compactly supported infinitely differentiable function $\Upsilon : \mathbb{C} \to \mathbb{C}$ with support disjoint from $\operatorname{supp} \mu_f$, we have for each $N > 0$ the representation $Y_N = nN\Re\mathbb{E}(\Upsilon(\mathbf{z})S_{\mu_f^N}(\mathbf{z})|\mathcal{F})$, almost surely. Furthermore, by similar reasoning, for any analytic function $b : \mathbb{C} \setminus \operatorname{supp} \mu_f \to \mathbb{C}$ satisfying $b(z^*) \equiv b(z)^*$, we have $\Re\mathbb{E}(\Upsilon(\mathbf{z})b(\mathbf{z})|\mathcal{F}) = 0$, almost surely. From statements (22) and (23) with $p = 4$ we deduce that $\sup_N N^{\frac{1}{2}}\|Y_{N+1} - Y_N\|_4 < \infty$ via Jensen's inequality in conditional form. From statements (24) and (25) we deduce that $\sup_N N\mathbb{E}Y_N < \infty$. Thus $Y_N \to_{N\to\infty} 0$, almost surely, by Lemma 2.5.3, which finishes the proof. □

2.6. **The self-adjoint linearization trick.** We now present a simple self-adjointness-preserving variant of the celebrated linearization trick of [10] and [9], and we explain how the trick gives access to the Stieltjes transforms $S_{\mu_f^N}(z)$ and $S_{\mu_f}(z)$ figuring in Theorem 2.4.6. This will motivate our focus on block-decomposed matrices and operators in the sequel.

2.6.1. *Schur complements.* Recall the familiar formula

$$\left[\begin{array}{cc} a & b \\ c & d \end{array}\right]^{-1} = \left[\begin{array}{cc} 0 & 0 \\ 0 & d^{-1} \end{array}\right] + \left[\begin{array}{c} 1 \\ -d^{-1}c \end{array}\right](a-bd^{-1}c)^{-1}\left[\begin{array}{cc} 1 & -bd^{-1} \end{array}\right] \tag{28}$$

for inverting a block-decomposed matrix. Formula (28) holds whenever d is invertible and at least one of $\left[\begin{array}{cc} a & b \\ c & d \end{array}\right]$ and the so-called *Schur complement* $a-bd^{-1}c$ are invertible, in which case both of the latter two matrices are invertible.

Proposition 2.6.2. *Fix $f \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X}\rangle)_{\mathrm{sa}}$ arbitrarily. Then for some integer $s > n$ there exists $\tilde{f} \in \mathrm{Mat}_s(\mathbb{C}\langle \mathbf{X}\rangle)_{\mathrm{sa}}$ all entries of which are of degree ≤ 1 in the variables $\mathbf{X}_\ell$ and which furthermore admits a block decomposition $\tilde{f} = \left[\begin{array}{cc} 0 & b \\ b^* & d \end{array}\right]$ such that $d \in \mathrm{Mat}_{s-n}(\mathbb{C}\langle \mathbf{X}\rangle)_{\mathrm{sa}}$ is invertible and $f = -bd^{-1}b^*$.*

We call $\tilde{f}$ a *self-adjoint linearization* of f. Later we will upgrade this definition to a slightly more sophisticated form. See Definition 5.4.3 below.

Proof. Let $\deg f$ be the maximum of the degrees of the noncommutative polynomials appearing in f. We proceed by induction on $\deg f$. There is no difficulty to find matrices $f_1, f_2 \in \mathrm{Mat}_{n\times N}(\mathbb{C}\langle \mathbf{X}\rangle)$ for some N such that $f = f_1 f_2^* + f_2 f_1^*$ and $\deg f_1 \vee \deg f_2 \leq 1 \vee \frac{2\deg f}{3}$. Put $b = \left[\begin{array}{cc} f_1 & f_2 \end{array}\right]$ and $d = -\left[\begin{array}{cc} 0 & \mathbf{I}_N \\ \mathbf{I}_N & 0 \end{array}\right]$. Note that $d = d^{-1}$ and $f = -bdb^*$. If $\deg f \leq 1$, then $g = \left[\begin{array}{cc} 0 & b \\ b^* & d \end{array}\right]$ is already a self-adjoint linearization of f and we are done. Otherwise, by induction on $\deg f$, the matrix g has a self-adjoint linearization $\tilde{g} = \left[\begin{array}{ccc} 0 & 0 & x \\ 0 & 0 & b_1 \\ x^* & b_1^* & d_1 \end{array}\right]$ where the zero block in the upper left is n-by-n, the central zero block is $2N$-by-$2N$ and the other blocks are of appropriate sizes. By (28) and the definitions we then have

$$\begin{aligned} \left[\begin{array}{cc} 0 & b \\ b^* & d \end{array}\right] &= -\left[\begin{array}{c} x \\ b_1 \end{array}\right] d_1^{-1}\left[\begin{array}{cc} x^* & b_1^* \end{array}\right], \\ & -\left[\begin{array}{cc} 0 & x \end{array}\right]\left[\begin{array}{cc} 0 & b_1 \\ b_1^* & d_1 \end{array}\right]^{-1}\left[\begin{array}{c} 0 \\ x^* \end{array}\right] \\ =& \quad -xd_1^{-1}x^* + xd_1^{-1}b_1^*(b_1d_1^{-1}b_1^*)^{-1}b_1d^{-1}x^* = -bd^{-1}b^* = f. \end{aligned}$$

Thus $\tilde{g}$ is a self-adjoint linearization of f as well as of g. □

Proposition 2.6.3. *Let $f \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X}\rangle)$ and $\tilde{f} = \left[\begin{array}{cc} 0 & b \\ b^* & d \end{array}\right] \in \mathrm{Mat}_s(\mathbb{C}\langle \mathbf{X}\rangle)$ be as in Proposition 2.6.2. Let $\mathcal{A}$ be any algebra, let $a = \{a_\ell\}_{\ell=1}^\infty$ be any sequence, fix $\Lambda \in \mathrm{Mat}_n(\mathcal{A})$ and let $\tilde{\Lambda} = \left[\begin{array}{cc} \Lambda & 0 \\ 0 & 0 \end{array}\right] \in \mathrm{Mat}_s(\mathcal{A})$. Then $f(a) - \Lambda$ is invertible if and only if $\tilde{f}(a) - \tilde{\Lambda}$ is invertible and under these equivalent conditions the upper left n-by-n block of $(\tilde{f}(a) - \tilde{\Lambda})^{-1}$ equals $(f(a) - \Lambda)^{-1}$.*

Proof. This follows immediately from (28) and the definitions. □

Remark 2.6.4. (In this remark we take for granted the notation and constructions of §3.4 and §4.1 below. There is no danger of circularity since we will not pick up

the discussion thread initiated here again until §5.) Proposition 2.6.3 says in the context of Theorem 2.4.6 that

$$S_{\mu_f^N}(z) \;=\; \frac{1}{Nn}\sum_{i=1}^{Nn}\left(\tilde f\left(\frac{\Xi^N}{\sqrt{N}}\right)-\left[\begin{array}{cc} z\mathbf{I}_{Nn} & 0\\ 0 & 0\end{array}\right]\right)^{-1}(i,i) \tag{29}$$

for $z \in \mathfrak{h}$ and combined with Proposition 3.4.5 below says furthermore that

$$\operatorname{supp}\mu_f = \operatorname{Spec}(f(\Xi)) \;=\; \left\{z\in\mathbb{C}\,\middle|\,\tilde f(\Xi)-\left[\begin{array}{cc} z\mathbf{I}_n & 0\\ 0 & 0\end{array}\right]\text{ not invertible}\right\}, \tag{30}$$

$$S_{\mu_f}(z) = \varphi^{\mathrm{BF}}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\tilde f(\Xi)-\left[\begin{array}{cc} z\mathbf{I}_n & 0\\ 0 & 0\end{array}\right]\right)^{-1}(i,i)\right) \tag{31}$$

for $z \in \mathbb{C}\setminus\operatorname{supp}\mu_f$. Now for $m \gg 0$ and suitable $a_0,\dots,a_m \in \operatorname{Mat}_s(\mathbb{C})_{\mathrm{sa}}$ we have

$$\tilde f = a_0\otimes 1_{\mathbb{C}\langle\mathbf{X}\rangle}+\sum_{\ell=1}^{m}a_\ell\otimes\mathbf{X}_\ell. \tag{32}$$

In order to gain control of the Stieltjes transforms $S_{\mu_f^N}(z)$ and $S_{\mu_f}(z)$, all our efforts hereafter will be directed toward understanding the special properties of block-decomposed random matrices of the form

$$\tilde f\left(\frac{\Xi^N}{\sqrt{N}}\right) = a_0\otimes\mathbf{I}_N+\sum_{\ell=1}^{m}a_\ell\otimes\frac{\Xi_\ell^N}{\sqrt{N}}\in\operatorname{Mat}_{sN}(\mathbb{C})_{\mathrm{sa}} \tag{33}$$

and of block-decomposed operators of the form

$$\tilde f(\Xi) = a_0\otimes 1_{B(\mathcal{H})}+\sum_{\ell=1}^{m}a_\ell\otimes\Xi_\ell\in\operatorname{Mat}_s(B(\mathcal{H}))_{\mathrm{sa}}. \tag{34}$$

Many tools from analysis and algebra will come into play.

3. Tools from operator theory

We review some topics in elementary C^*-algebra theory. We emphasize the viewpoint of noncommutative probability. We recall noncommutative laws, relations between laws and spectra, the construction of Boltzmann-Fock space, and Schur complements in the C^*-algebra context. Finally, we solve an abstract algebraic version of the Schwinger-Dyson equation.

3.1. **Warmup exercises.** We record without proof some very elementary facts frequently used below. Recall that we only use algebras $\mathcal{A}$ possessing a unit $1_{\mathcal{A}}$.

Lemma 3.1.1. *Let x and y be elements of a Banach algebra with x invertible and $2\left[\!\left[x^{-1}\right]\!\right]\left[\!\left[y\right]\!\right]\le 1$. Then $x-y$ is invertible and $\left[\!\left[(x-y)^{-1}\right]\!\right]\le 2\left[\!\left[x^{-1}\right]\!\right]$.*

Here and below we invariably use $[\![\cdot]\!]$ to denote the norm on a Banach algebra.

3.1.2. We note the *resolvent identity*

$$x^{-1}-y^{-1} = y^{-1}(y-x)x^{-1} = x^{-1}(y-x)y^{-1}\quad(x,y\in\mathcal{A}^\times) \tag{35}$$

holding in any algebra $\mathcal{A}$ and its infinitesimal variant $\frac{d}{dt}x^{-1} = -x^{-1}\frac{dx}{dt}x^{-1}$. We also need the iterated version

$$x^{-1}-y^{-1} = y^{-1}(y-x)y^{-1}+y^{-1}(y-x)y^{-1}(y-x)x^{-1}\quad(x,y\in\mathcal{A}^\times). \tag{36}$$

3.2. **Positivity.** We recall basic facts about positive elements of C^*-algebras.

3.2.1. *Positive elements and their square roots.* If an element x of a C^*-algebra $\mathcal{A}$ is self-adjoint with nonnegative spectrum, we write $x \geq 0$; and if furthermore x is invertible, then we write $x > 0$. Elements satisfying $x \geq 0$ are called *positive.* Elements of the form xx^* are automatically positive. For $x \in \mathcal{A}$ such that $x \geq 0$, there exists unique $y \in \mathcal{A}$ such that $y \geq 0$ and $y^2 = x$ (see [17, Thm. 2.2.1]), in which case we write $x^{1/2} = y$.

3.2.2. *C^*-subalgebras and GNS.* Let $\mathcal{A}$ be a C^*-algebra. We say that a closed subspace $\mathcal{A}_0 \subset \mathcal{A}$ is a *C^*-subalgebra* if $\mathcal{A}_0$ is stable under $*$, closed under multiplication and furthermore $1_\mathcal{A} \in \mathcal{A}_0$, in which case $\mathcal{A}_0$ is a C^*-algebra in its own right for which $1_{\mathcal{A}_0} = 1_\mathcal{A}$. Each C^*-algebra is isomorphic to a C^*-subalgebra of $B(H)$ for some Hilbert space H via the GNS construction (see [17, §3.4]).

Proposition 3.2.3. *For any C^*-algebra $\mathcal{A}$ and C^*-subalgebra $\mathcal{A}_0 \subset \mathcal{A}$ we have $\mathcal{A}_0 \cap \mathcal{A}^\times = \mathcal{A}_0^\times$.*

(See [17, Thm. 2.1.11].) Thus the spectrum of $x \in \mathcal{A}_0$ is the same whether viewed in $\mathcal{A}_0$ or $\mathcal{A}$. In particular, x is positive in $\mathcal{A}_0$ if and only if positive in $\mathcal{A}$.

Proposition 3.2.4. *For every element x of a C^*-algebra $\mathcal{A}$, if x is normal, and in particular, if x is self-adjoint, then $[\![x]\!]$ equals the spectral radius of x. Consequently, $[\![x]\!]^2$ equals the spectral radius of xx^* and x^*x.*

(See [17, Thm. 2.1.1 and Cor. 2.1.2].) It follows that a $*$-algebra can be normed as C^*-algebra in at most one way. We always use that norm when it exists.

3.2.5. *Real and imaginary parts.* Given any $*$-algebra and $Z \in \mathcal{A}$ we write $\Re Z = \frac{Z+Z^*}{2}$ and $\Im Z = \frac{Z-Z^*}{2\mathrm{i}}$. (This generalizes the notation we already have for real and imaginary parts of a complex number.)

The next elementary result plays an vitally important role in the paper.

Lemma 3.2.6. *Let $\mathcal{A}$ be a C^*-algebra. Let $A \in \mathcal{A}$ satisfy $\Im A \geq 0$ and let $z \in \mathfrak{h}$. Then $A + z1_\mathcal{A} \in \mathcal{A}^\times$ and $\left[\!\left[(A + z1_\mathcal{A})^{-1}\right]\!\right] \leq 1/\Im z$.*

Proof. To abbreviate we write $1 = 1_\mathcal{A}$, $z = z1_\mathcal{A}$, and so on. After replacing A by $(A+\Re z)/\Im z$ we may assume without loss of generality that $z = \mathrm{i}$. Write $A = X+\mathrm{i}Y$ with $X = \Re A$ and $Y = \Im A$. Since $Y \geq 0$, we have $1+Y > 0$, and hence we can write $A + \mathrm{i} = (1 + Y)^{1/2}(W + \mathrm{i})(1 + Y)^{1/2}$ where $W = (1 + Y)^{-1/2}X(1 + Y)^{-1/2} \in \mathcal{A}_{\mathrm{sa}}$. Since both $(1 + Y)^{1/2}$ and $W + \mathrm{i}$ are normal and have spectra disjoint from the open unit disc centered at the origin, both are invertible with inverse of norm ≤ 1 by Proposition 3.2.4. Thus $A + \mathrm{i}$ is invertible with inverse of norm ≤ 1. □

3.3. **States and spectral theory.** We recall some basic definitions and results pertaining to C^*-probability spaces. Much of this background is covered in [17]. The rest of it is more or less implicit in [17] and [21] but hard to extract. Some of this material is also covered in [1, Chap. 1] but unfortunately Lemma 3.3.6 below is not. For the reader's convenience we supply short proofs of some key statements which are part of standard "C^*-know-how" but hard to pin down in the literature.

3.3.1. *States.* Let $\mathcal{A}$ be a C^*-algebra. Let $\phi : \mathcal{A} \to \mathbb{C}$ be any linear functional (perhaps not bounded). One calls ϕ *positive* if for every $A \in \mathcal{A}$, if $A \geq 0$, then $\phi(A) \geq 0$, in which case ϕ is automatically bounded and satisfies $\phi(x^*) = \phi(x)^*$. One calls ϕ a *state* if ϕ is positive and $\phi(1_\mathcal{A}) = 1$, in which case $[\![\phi]\!] = 1$. One calls

a state ϕ *faithful* if for every $A \in \mathcal{A}$, if $A \geq 0$ and $A \neq 0$, then $\phi(A) > 0$. Note that by Proposition 3.2.3, for any C^*-subalgebra $\mathcal{A}_0 \subset \mathcal{A}$ and state ϕ on $\mathcal{A}$ the restriction of ϕ to $\mathcal{A}_0$ is again a state. (All of this is covered in [17, Chap. 3].)

Definition 3.3.2. A pair $(\mathcal{A}, \phi)$ consisting of a C^*-algebra $\mathcal{A}$ and a state ϕ is called a *C^*-probability space*. We call $(\mathcal{A}, \phi)$ *faithful* if ϕ is so.

3.3.3. *Laws of noncommutative random variables.* Given a C^*-probability space $(\mathcal{A}, \phi)$ and self-adjoint $A \in \mathcal{A}$, there exists a unique Borel probability measure μ_A on the spectrum of A, called the *law* of A, such that $\phi(f(A)) = \int f d\mu_A$ for every continuous $\mathbb{C}$-valued function f on the spectrum of A, where $f(A)$ is defined by means of the functional calculus at A, i.e., the inverse Gelfand transform, and μ_A is provided by the Riesz representation theorem. For convenience we always extend the law μ_A to a Borel probability measure on the real line supported on $\mathrm{Spec}(A)$. (See [1, Chap. 5] for background on laws.) We note the important formula

$$S_{\mu_A}(z) = \phi((A - z1_{\mathcal{A}})^{-1}) \tag{37}$$

for the Stieltjes transform of the law μ_A which holds for every $z \in \mathbb{C}$ belonging neither to the support of μ_A nor to the spectrum of A. A simple and useful criterion for equality of the latter two sets is provided by the next result.

Lemma 3.3.4. *Let $(\mathcal{A}, \phi)$ be a faithful C^*-probability space. Then, for every $A \in \mathcal{A}_{\mathrm{sa}}$, $\mathrm{supp}\, \mu_A = \mathrm{Spec}(A)$.*

Proof. Let $K = \mathrm{Spec}(A) \subset \mathbb{R}$, noting that K is compact. Let $\mathcal{A}_0 \subset \mathcal{A}$ be the C^*-subalgebra generated by A and put $\phi_0 = \phi|_{\mathcal{A}_0}$, which is a faithful state on $\mathcal{A}_0$. By the theory of the Gelfand transform, $\mathcal{A}_0$ can be identified with the C^*-algebra of continuous complex-valued functions defined on K. Under this identification the operator A becomes the identity function $\mathrm{Spec}(A) \to \mathbb{R}$ and ϕ_0 becomes the linear functional represented by μ_A. By Urysohn's Lemma, ϕ_0 cannot be faithful unless $\mathrm{supp}\, \mu_A = K$. □

Lemma 3.3.5. *If $(\mathcal{A}, \phi)$ is a faithful C^*-probability space, then so is $(\mathrm{Mat}_n(\mathcal{A}), \phi_n)$, where $\phi_n(A) = \frac{1}{n}\sum_{i=1}^n \phi(A(i,i))$.*

We always follow the procedure of this lemma to equip $\mathrm{Mat}_n(\mathcal{A})$ with a state when given a state on $\mathcal{A}$.

Proof. There is exactly one way to norm the $*$-algebra $\mathrm{Mat}_n(\mathcal{A})$ as a C^*-algebra. (See [17, Thm. 3.4.2] and also §4.1.6 below.) Following our convention to norm every $*$-algebra as a C^*-algebra when possible, we thus regard $\mathrm{Mat}_n(\mathcal{A})$ as a C^*-algebra. For $0 \neq A \in \mathrm{Mat}_n(\mathcal{A})$ such that $A \geq 0$,

$$\phi_n(A) = \phi_n(A^{1/2}A^{1/2}) = \frac{1}{n}\sum_{i,j=1}^n \phi(A^{1/2}(i,j)A^{1/2}(i,j)^*).$$

This formula first of all make it clear that ϕ_n is a state and hence that $(\mathrm{Mat}_n(\mathcal{A}), \phi_n)$ is a C^*-probability space. But furthermore, at least one term on the right is > 0 since ϕ is faithful and $A^{1/2} \neq 0$. Thus ϕ_n is faithful. □

Lemma 3.3.6. *Let H be a Hilbert space, let $v \in H$ be a unit vector, and consider the vectorial state $\phi = (A \mapsto (v, Av)) : B(H) \to \mathbb{C}$ associated with v. Let $\mathcal{A}, \hat{\mathcal{A}} \subset B(H)$ be C^*-subalgebras such that $A\hat{A} = \hat{A}A$ for all $A \in \mathcal{A}$ and $\hat{A} \in \hat{\mathcal{A}}$.*

Assume furthermore that the vector v is cyclic *for $\hat{\mathcal{A}}$, i.e., that the set $\{\hat{A}v \mid \hat{A} \in \hat{\mathcal{A}}\}$ is dense in H. Then $\phi|_{\mathcal{A}}$ is faithful.*

Proof. Fix $A \in \mathcal{A}$ such that $A \geq 0$ and $A \neq 0$. Clearly there exists $h \in H$ such that $(h, A^{1/2}h) > 0$. Thus by hypothesis there exists $\hat{A} \in \hat{\mathcal{A}}$ such that

$$0 < (\hat{A}v, A^{1/2}\hat{A}v) = (v, \hat{A}^* A^{1/2}\hat{A}v) = \phi(\hat{A}^* A^{1/2}\hat{A}) = \phi(\hat{A}^*\hat{A}A^{1/2}).$$

Making further use of the hypothesis that operators in $\mathcal{A}$ commute with operators in $\hat{\mathcal{A}}$, we have

$$0 \leq \phi((\sqrt{t}\hat{A}^*\hat{A} - A^{1/2}/\sqrt{t})^2) = t\phi((\hat{A}^*\hat{A})^2) + \phi(A)/t - 2\phi(\hat{A}^*\hat{A}A^{1/2})$$

for $t > 0$. The last inequality forces $\phi(A) > 0$. Thus $\phi|_{\mathcal{A}}$ is indeed faithful. □

3.4. **Boltzmann-Fock space.** We now recall the standard construction of free semicircular variables laying stress on some properties of the construction which are not often exploited in practice, but which will be important here. We are essentially just summarizing (using somewhat different notation) enough material from [21] to take advantage of [21, Remark 2.6.7].

3.4.1. *Definition of $\mathcal{H}$ and the C^*-probability space $(B(\mathcal{H}), \phi)$.* Let $\mathcal{H}$ be a Hilbert space canonically equipped with an orthonormal basis $\{v(i_1 \cdots i_k)\}$ indexed by all finite sequences of positive integers, including the empty sequence. We write $1_{\mathcal{H}} = v(\emptyset) \in \mathcal{H}$. We equip $B(\mathcal{H})$ with the vectorial state ϕ^{BF} defined by

$$\phi^{\mathrm{BF}}(A) = (1_{\mathcal{H}}, A1_{\mathcal{H}}),$$

thus making it into a noncommutative probability space. (Note that we take Hilbert space inner products to be linear on the right, antilinear on the left.) Context permitting, we drop the superscript and write $\phi = \phi^{\mathrm{BF}}$.

3.4.2. *Raising and lowering operators.* Let $\Sigma_i \in B(\mathcal{H})$ act by the rule

$$\Sigma_i v(i_1 \cdots i_k) = v(ii_1 \cdots i_k).$$

Let $p_{\mathcal{H}} \in B(\mathcal{H})$ denote orthogonal projection to the linear span of $1_{\mathcal{H}}$. It is easy to verify the following relations, where i and j are any positive integers:

$$p_{\mathcal{H}}\Sigma_i = 0 = \Sigma_i^* p_{\mathcal{H}}, \ \ \Sigma_i^*\Sigma_j = \delta_{ij}1_{B(\mathcal{H})}, \ \ [\![\Sigma_i]\!] = [\![\Sigma_i^*]\!] = 1, \tag{38}$$

$$\phi(\Sigma_i) = \phi(\Sigma_i^*) = 0, \ \ \phi(\Sigma_i\Sigma_j) = \phi(\Sigma_i\Sigma_j^*) = \phi(\Sigma_i^*\Sigma_j^*) = 0, \ \ \phi(\Sigma_i^*\Sigma_j) = \delta_{ij}. \tag{39}$$

3.4.3. *The semicircular variables Ξ_ℓ.* We now make the special choice of sequence Ξ mentioned in Remark 2.4.7 above. Put

$$\Xi_\ell = \mathrm{i}^\ell \Sigma_\ell + \mathrm{i}^{-\ell}\Sigma_\ell^*$$

for all ℓ and put $\Xi = \{\Xi_\ell\}_{\ell=1}^\infty \in B(\mathcal{H})^\infty$. It is well known that the joint law of the sequence $\{\Sigma_\ell + \Sigma_\ell^*\}_{\ell=1}^\infty$ is free semicircular. See [21] or [1, Chap. 5]. It is easy to see that the sequences Ξ and $\{\Sigma_\ell + \Sigma_\ell^*\}_{\ell=1}^\infty$ have the same joint law. Indeed, the former is conjugate to the latter by a unitary transformation preserving ϕ^{BF}. We work exclusively with the choice of Ξ made in this paragraph for the rest of the paper.

3.4.4. *Right raising and lowering operators.* For each integer $i > 0$, let $\hat{\Sigma}_i \in B(\mathcal{H})$ act on $\mathcal{H}$ by the rule

$$\hat{\Sigma}_i v(i_1 \cdots i_k) = v(i_1 \cdots i_k i).$$

In direct analogy to (38) we have

$$p_{\mathcal{H}} \hat{\Sigma}_i = 0 = \hat{\Sigma}_i^* p_{\mathcal{H}}, \quad \hat{\Sigma}_i^* \hat{\Sigma}_j = \delta_{ij} 1_{B(\mathcal{H})}, \quad \left[\!\left[\hat{\Sigma}_i\right]\!\right] = \left[\!\left[\hat{\Sigma}_i^*\right]\!\right] = 1. \tag{40}$$

We also have right analogues of the relations (39) but we will not need them. It is easy to verify the following relations, where i and j are any positive integers:

$$\hat{\Sigma}_i \Sigma_j = \Sigma_j \hat{\Sigma}_i, \quad \hat{\Sigma}_j^* \Sigma_i = \Sigma_i \hat{\Sigma}_j^* + \delta_{ij} p_{\mathcal{H}}, \quad \Sigma_i p_{\mathcal{H}} = \hat{\Sigma}_i p_{\mathcal{H}}. \tag{41}$$

It is trivial but important to note that every relation above implies another by taking adjoints on both sides.

Proposition 3.4.5. *For all $f \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X} \rangle)_{\mathrm{sa}}$, $\mathrm{supp}\,\mu_f = \mathrm{Spec}(f(\Xi))$.*

Proof. Let $\mathcal{A} \subset B(\mathcal{H})$ be the C^*-subalgebra generated by the sequence Ξ. By Lemmas 3.3.4 and 3.3.5, it is enough to show that $\phi|_{\mathcal{A}}$ is faithful. Let

$$\hat{\Xi}_\ell = \mathrm{i}^\ell \hat{\Sigma}_\ell + \mathrm{i}^{-\ell} \hat{\Sigma}_\ell^* \in B(\mathcal{H})_{\mathrm{sa}}$$

for positive integers ℓ and let $\hat{\mathcal{A}} \subset B(\mathcal{H})$ be the C^*-subalgebra of $B(\mathcal{H})$ generated by the sequence $\{\hat{\Xi}_\ell\}$. Using (41), one verifies that

$$\Xi_\ell \hat{\Xi}_m = \hat{\Xi}_m \Xi_\ell$$

for all ℓ and m. Here we are using the powerful insight of [21, Remark 2.6.7]. Thus every element of $\mathcal{A}$ commutes with every element of $\hat{\mathcal{A}}$. It is also easy to see that $1_{\mathcal{H}}$ is cyclic for $\hat{\mathcal{A}}$. Therefore $\phi|_{\mathcal{A}}$ is faithful by Lemma 3.3.6. □

We conclude our discussion of Boltzmann-Fock space by recording the following "hypothesis-checking" result for use in §6.

Lemma 3.4.6. *Fix a positive integer m. Let*

$$x \in \{1_{B(\mathcal{H})}\} \cup \{\Sigma_j, \Sigma_j^* \mid j = 1, \ldots, m\}.$$

The following hold:

- $\hat{\Sigma}_i^* x \hat{\Sigma}_j = \delta_{ij} x$ *for all i and j.*
- $p_{\mathcal{H}} x \hat{\Sigma}_i = p_{\mathcal{H}} x \hat{\Sigma}_i p_{\mathcal{H}}$ *and* $p_{\mathcal{H}} \hat{\Sigma}_i^* x p_{\mathcal{H}} = \hat{\Sigma}_i^* x p_{\mathcal{H}}$ *for all i.*
- *x commutes with $\hat{\Sigma}_i$ and $\hat{\Sigma}_i^*$ for all $i > m$.*
- *x commutes with $p_{\mathcal{H}} + \sum_{i=1}^m \hat{\Sigma}_i \hat{\Sigma}_i^*$.*

Proof. The first three statements follow straightforwardly from (38), (40) and (41), so we just supply a proof for the last statement. We write $[A, B] = AB - BA$. Note that $[A, BC] = [A, B]C + B[A, C]$. Fix $j \in \{1, \ldots, m\}$. We then have

$$\begin{aligned} & [\Sigma_j, p_{\mathcal{H}} + \textstyle\sum_{i=1}^m \hat{\Sigma}_i \hat{\Sigma}_i^*] = [\Sigma_j, p_{\mathcal{H}}] + \textstyle\sum_{i=1}^m ([\Sigma_j, \hat{\Sigma}_i] \hat{\Sigma}_i^* + \hat{\Sigma}_i [\Sigma_j, \hat{\Sigma}_i^*]) \\ = \; & \Sigma_j p_{\mathcal{H}} + \hat{\Sigma}_j [\Sigma_j, \hat{\Sigma}_j^*] = \Sigma_j p_{\mathcal{H}} - \hat{\Sigma}_j p_{\mathcal{H}} = 0. \end{aligned}$$

The analogous relation with Σ_j^* in place of Σ_j follows by taking adjoints. □

3.5. **Projections, inverses and Schur complements.** We make an *ad hoc* extension of the Schur complement concept to the context of C^*-algebras.

3.5.1. *Projections and π-inverses.* Let $\mathcal{A}$ be a C^*-algebra. A *projection* $\pi \in \mathcal{A}$ by definition satisfies $\pi = \pi^* = \pi^2$. A family $\{\pi_i\}$ of projections is called *orthonormal* if $\pi_i \neq 0$ and $\pi_i\pi_j = \delta_{ij}\pi_i$ for all i and j. Given $x \in \mathcal{A}$ and a projection $0 \neq \pi \in \mathcal{A}$, we denote by x_π^{-1} the inverse of $\pi x \pi$ in the C^*-algebra $\pi\mathcal{A}\pi$, if it exists, in which case it is uniquely defined. We call x_π^{-1} the *π-inverse* of x. Note that $x_\pi^{-1} = (\pi x \pi)_\pi^{-1}$.

Proposition 3.5.2. *Let $\mathcal{A}$ be a C^*-algebra. Let $\{\pi, \pi^\perp\}$ be an orthonormal system of projections in $\mathcal{A}$ and put $\sigma = \pi + \pi^\perp$. Let $x \in \mathcal{A}$ satisfy $\pi^\perp x \pi^\perp \in (\pi^\perp \mathcal{A} \pi^\perp)^\times$. Then we have*

$$\sigma x \sigma \in (\sigma\mathcal{A}\sigma)^\times \Leftrightarrow \pi(x - x x_{\pi^\perp}^{-1} x)\pi \in (\pi\mathcal{A}\pi)^\times \tag{42}$$

and under these equivalent conditions we have

$$\pi x_\sigma^{-1} \pi = (x - x x_{\pi^\perp}^{-1} x)_\pi^{-1}, \tag{43}$$

$$x_\sigma^{-1} - x_{\pi^\perp}^{-1} = (\pi - x_{\pi^\perp}^{-1} x \pi) x_\sigma^{-1} (\pi - \pi x x_{\pi^\perp}^{-1}). \tag{44}$$

Thus the expression $\pi(x - x x_{\pi^\perp}^{-1} x)\pi$ is a sort of generalized Schur complement.

Proof. By exploiting the GNS construction one can interpret the proposition as an instance of (28). We omit the details. □

3.6. **Cuntz frames and quasi-circularity.** We elaborate upon a suggestion made in the last exercise of [1, Chap. 5]. We fix a C^*-algebra $\mathcal{A}$.

Definition 3.6.1. Suppose we are given a collection $\{\pi\} \cup \{\rho_i\}_{i=1}^\infty$ of elements of $\mathcal{A}$ satisfying the following conditions:

$$\pi \text{ is a nonzero projection, } \pi\rho_i = 0 \text{ and } \rho_i^*\rho_j = \delta_{ij} 1_{\mathcal{A}} \text{ for all } i \text{ and } j. \tag{45}$$

We call $\{\pi\} \cup \{\rho_i\}_{i=1}^\infty$ a *Cuntz frame* in $\mathcal{A}$. Note that $\{\pi\} \cup \{\rho_i\rho_i^*\}_{i=1}^\infty$ is automatically an orthonormal system of projections.

Remark 3.6.2. The relations $\rho_i^*\rho_j = \delta_{ij} 1_{\mathcal{A}}$ are those defining the *Cuntz algebra* [6], hence our choice of terminology.

3.6.3. *Quasi-circular operators.* Suppose we are given a Cuntz frame $\{\pi\} \cup \{\rho_i\}_{i=1}^\infty$ in $\mathcal{A}$. We say that an operator $A \in \mathcal{A}$ is *quasi-circular* (with respect to the given Cuntz frame) if the following statements hold:

$$\rho_i^* A \rho_j = \delta_{ij} A \text{ for all } i \text{ and } j. \tag{46}$$

$$\pi A \rho_i \pi = \pi A \rho_i \text{ and } \pi\rho_i^* A \pi = \rho_i^* A \pi \text{ for all } i. \tag{47}$$

(48) There exists an integer $k_A \geq 0$ such that A commutes with $\pi + \sum_{i=1}^{k_A} \rho_i\rho_i^*$ and also with ρ_i and ρ_i^* for all $i > k_A$.

Proposition 3.6.4. *Let $\{\pi\} \cup \{\rho_i\}_{i=1}^\infty$ be a Cuntz frame in $\mathcal{A}$. Let $A \in \mathcal{A}^\times$ be quasi-circular with respect to the given frame. Choose any integer $k \geq k_A$. Then*

$$\pi A^{-1} \pi = \left(\pi A \pi - \sum_{i=1}^k \pi A \rho_i \pi A^{-1} \pi \rho_i^* A \pi\right)_\pi^{-1}. \tag{49}$$

In particular, one automatically has $\pi A^{-1}\pi \in (\pi\mathcal{A}\pi)^\times$.

Identity (49) is an abstract algebraic version of the Schwinger-Dyson equation. See the proof of Proposition 6.2.2 below for the application.

Proof. Consider the projections $\sigma = \pi + \sum_{i=1}^k \rho_i\rho_i^*$ and $\pi^\perp = \sigma - \pi$. We claim that

$$A_{\pi^\perp}^{-1} = \textstyle\sum_{i=1}^k \rho_i A^{-1}\rho_i^*. \tag{50}$$

In any case, we have $\pi^\perp A \pi^\perp = \sum_{i=1}^k \rho_i A \rho_i^*$ by (46). Furthermore, we have

$$(\textstyle\sum_{i=1}^k \rho_i A\rho_i^*)(\sum_{j=1}^k \rho_j A^{-1}\rho_j^*) = \pi^\perp = (\sum_{i=1}^k \rho_i A^{-1}\rho_i^*)(\sum_{j=1}^k \rho_j A\rho_j^*)$$

by (45). Thus claim (50) holds. To prove (49), we calculate as follows:

$$\begin{array}{rcl} \pi A^{-1}\pi &=& \pi\sigma A^{-1}\sigma\pi \;=\; \pi A_\sigma^{-1}\pi \;=\; (A - AA_{\pi^\perp}^{-1}A)_\pi^{-1} \\ &=& (\pi A\pi - \pi AA_{\pi^\perp}^{-1}A\pi)_\pi^{-1} = (\pi A\pi - \sum_{i=1}^k \pi A\rho_i A^{-1}\rho_i^* A\pi)_\pi^{-1} \\ &=& (\pi A\pi - \sum_{i=1}^k \pi A\rho_i\pi A^{-1}\pi\rho_i^* A\pi)_\pi^{-1}. \end{array}$$

The first step is simply an exploitation of orthonormality of $\{\pi, \pi^\perp\}$. Since A commutes with σ by (48), we have $\sigma A^{-1}\sigma A = \sigma = 1_{\sigma A\sigma} = A\sigma A^{-1}\sigma$, which justifies the second step. The third step is an application of (43) and the fourth step is a trivial consequence of the definition of π-inverse. The fifth step is an application of (50) and the last step is an application of (47). The proof of (49) is complete. □

Remark 3.6.5. The preceding calculation will obviate consideration of combinatorics of free semicircular variables in the sequel. We present this approach as counterpoint to the nowadays standard combinatorial approach discussed briefly in [1, Chap. 5] and developed at length in [18].

4. Tools for concentration

In this section we introduce an ensemble of tools we use to (partially) replace the Poincaré-type inequalities used in [10], [9], [20], [5] and [15]. We speak of an ensemble because no one tool seems to contribute more than incrementally.

4.1. **Tensor products and norming rules.**
We rehearse the most basic rules of calculation and estimation used in the paper.

4.1.1. *Tensor products of vector spaces and algebras.* Given vector spaces $\mathcal{A}$ and $\mathcal{B}$ over $\mathbb{C}$, let $\mathcal{A} \otimes \mathcal{B}$ denote the tensor product of $\mathcal{A}$ and $\mathcal{B}$ formed over $\mathbb{C}$. If $\mathcal{A}$ and $\mathcal{B}$ are both algebras, we invariably endow $\mathcal{A} \otimes \mathcal{B}$ with algebra structure by the rule $(a_1 \otimes b_1)(a_2 \otimes b_2) = a_1a_2 \otimes b_1b_2$. If $\mathcal{A}$ and $\mathcal{B}$ are both $*$-algebras, we invariably endow $\mathcal{A} \otimes \mathcal{B}$ with $*$-algebra structure by the rule $(a \otimes b)^* = a^* \otimes b^*$.

4.1.2. *Tensor notation for building matrices.* Let $\mathcal{A}$ be an algebra. We identify the algebra $\mathrm{Mat}_n(\mathbb{C}) \otimes \mathcal{A}$ with $\mathrm{Mat}_n(\mathcal{A})$ by the rule $(X \otimes a)(i,j) = x(i,j)a$ and more generally use the same rule to identify the space $\mathrm{Mat}_{k\times\ell}(\mathbb{C}) \otimes \mathcal{A}$ with the space of rectangular matrices $\mathrm{Mat}_{k\times\ell}(\mathcal{A})$. Furthermore, in the case $\mathcal{A} = \mathrm{Mat}_s(\mathbb{C})$, we identify $X \otimes a$ with an element of $\mathrm{Mat}_{ks\times\ell s}(\mathbb{C})$ by viewing $X \otimes a$ as a k-by-ℓ arrangement of s-by-s blocks $X(i,j)a$. In other words, we identify $X \otimes a$ with the usual Kronecker product of X and a.

4.1.3. *Banach spaces.* Banach spaces always have $\mathbb{C}$ as scalar field, and bounded (multi)linear maps between Banach spaces are always $\mathbb{C}$-(multi)linear, unless explicitly noted otherwise. To avoid collision with the notation $\|\cdot\|_p$, we let $[\![\cdot]\!]_{\mathcal{V}}$ denote the norm of a Banach space $\mathcal{V}$ and context permitting (nearly always), we drop the subscript.

4.1.4. *(Multi)linear maps between Banach spaces.* Given Banach spaces $\mathcal{V}$ and $\mathcal{W}$, let $B(\mathcal{V},\mathcal{W})$ denote the space of bounded linear maps $\mathcal{V}\to\mathcal{W}$. Let $B(\mathcal{V})=B(\mathcal{V},\mathcal{V})$ and let $\mathcal{V}^\star$ denote the linear dual of $\mathcal{V}$. Given $T\in B(\mathcal{V},\mathcal{W})$, let $[\![T]\!]=[\![T]\!]_{B(\mathcal{V},\mathcal{W})}$ be the best constant such that $[\![Tv]\!]\leq[\![T]\!]\,[\![v]\!]$. We always use the norm on $B(\mathcal{V},\mathcal{W})$ so defined. More generally, let $B(\mathcal{V}_1,\dots,\mathcal{V}_r;\mathcal{W})$ denote the space of bounded r-linear maps $\mathcal{V}_1\times\cdots\times\mathcal{V}_r\to\mathcal{W}$ and given $T\in B(\mathcal{V}_1,\dots,\mathcal{V}_r;\mathcal{W})$, let $[\![T]\!]=[\![T]\!]_{B(\mathcal{V}_1,\dots,\mathcal{V}_r;\mathcal{W})}$ be the best constant such that $[\![T(v_1,\dots,v_r)]\!]\leq[\![T]\!]\,[\![v_1]\!]\cdots[\![v_r]\!]$. We always use the norm on $B(\mathcal{V}_1,\dots,\mathcal{V}_r;\mathcal{W})$ so defined.

4.1.5. *Matrix spaces over C^*-algebras.* Let $\mathcal{A}$ be any C^*-algebra. We have already noted in the proof of Lemma 3.3.5 that there is a unique way to norm the $*$-algebra $\mathrm{Mat}_n(\mathcal{A})$ as a C^*-algebra. In turn, we always norm the space of rectangular matrices $\mathrm{Mat}_{k\times\ell}(\mathcal{A})$ by the formula $[\![A]\!]=[\![AA^*]\!]^{1/2}$. Note that

$$\bigvee_{i=1}^{k}\bigvee_{j=1}^{\ell}[\![A(i,j)]\!]\leq[\![A]\!]\leq\sum_{m=-\infty}^{\infty}\bigvee_{i=1}^{k}\bigvee_{j=1}^{\ell}[\![A(i,j)]\!]\,\mathbf{1}_{i-j=m}. \tag{51}$$

Moreover, given $B\in\mathrm{Mat}_{\ell\times m}(\mathcal{A})$, we have $[\![AB]\!]\leq[\![A]\!]\,[\![B]\!]$. In particular, for every square or rectangular matrix A with complex number entries, $[\![A]\!]$ is the largest singular value of A.

4.1.6. *Tensor products of C^*-algebras.* Given C^*-algebras $\mathcal{A}$ and $\mathcal{B}$ with at least one of them finite-dimensional, the $*$-algebra $\mathcal{A}\otimes\mathcal{B}$ has exactly one C^*-algebra norm. To see this, only existence requires comment since uniqueness we have already noted after Proposition 3.2.4. We proceed as follows. Firstly, we observe that since $\mathcal{A}\otimes\mathcal{B}$ and $\mathcal{B}\otimes\mathcal{A}$ are isomorphic $*$-algebras, we may assume that $\mathcal{A}$ is finite-dimensional. Then, after reducing to the case $\mathcal{A}=\mathrm{Mat}_n(\mathbb{C})$ and $\mathcal{B}=B(H)$ by using the GNS construction, we can make identifications $\mathcal{A}\otimes\mathcal{B}=\mathrm{Mat}_n(B(H))=B(H^n)$ yielding the desired norm. Thus existence is settled. The preceding argument shows that for all $a\in\mathcal{A}$ and $b\in\mathcal{B}$ we have $[\![a\otimes b]\!]=[\![a]\!]\,[\![b]\!]$. In a similar vein we have the following useful general observation.

Lemma 4.1.7. *Let $\mathcal{S}$ be a finite-dimensional C^*-algebra. Let $\{e_i\}_{i=1}^n$ be any linearly independent family of elements of $\mathcal{S}$. Then for all C^*-algebras $\mathcal{A}$ and families $\{a_i\}_{i=1}^n$ of elements of $\mathcal{A}$ we have*

$$\frac{1}{C}\bigvee_{i=1}^{n}[\![a_i]\!]\leq\left[\!\!\left[\sum_{i=1}^{n}e_i\otimes a_i\right]\!\!\right]=\left[\!\!\left[\sum_{i=1}^{n}a_i\otimes e_i\right]\!\!\right]\leq C\sum_{i=1}^{n}[\![a_i]\!]$$

for a constant $C\geq 1$ depending only on $\mathcal{S}$ and $\{e_i\}$.

Proof. We may assume that $\mathcal{S}=\mathrm{Mat}_s(\mathbb{C})$. Furthermore, there is no loss of generality to assume that $n=s^2$ and thus that $\{e_i\}_{i=1}^{s^2}$ is a basis for $\mathrm{Mat}_s(\mathbb{C})$. Finally, there is no loss of generality to assume that $\{e_i\}_{i=1}^{s^2}$ consists of elementary matrices, in which case the lemma at hand reduces to (51). $\square$

4.1.8. *Block algebras.* We define a *block algebra* to be a C^*-algebra isomorphic to $\mathrm{Mat}_s(\mathbb{C})$ for some positive integer s. The point of the definition is merely to compress notation and to put some psychological distance between us and the too-familiar algebra $\mathrm{Mat}_s(\mathbb{C})$. (Later we will refine the notion of block algebra by keeping track not only of the transpose conjugate operation but also of the transpose. See §5.2 below.)

4.2. **Quadratic forms in independent random vectors.** Variants of the next result are in common use in RMT. (See, e.g., [3, Lemma 2.7].)

Proposition 4.2.1. *Let $Y_1, \dots, Y_n$ and $Z_1, \dots, Z_n$ be $\mathbb{C}$-valued random variables which for some $p \in [2, \infty)$ all belong to L^{2p} and have mean zero. Let $A \in \mathrm{Mat}_n(\mathbb{C})$ be a (deterministic) matrix. Assume furthermore that the family of σ-fields $\{\sigma(Y_i, Z_i)\}_{i=1}^n$ is independent. Then we have*

$$\left\| \sum_{i,j=1}^n A(i,j)(Y_i Z_j - \mathbb{E} Y_i Z_j) \right\|_p \leq c \left(\sum_{i,j=1}^n |A(i,j)|^2 \|Y_i\|_{2p}^2 \|Z_j\|_{2p}^2 \right)^{1/2}$$

for a constant c depending only on p.

Results of this type are well-known. The earliest reference we know is [22]. In that reference the result above is proved in the special case in which $Y_i = Z_i = Y_i^* = Z_i^*$ and A has real entries. From that special case the general case of the proposition above can be deduced easily by algebraic manipulation.

We now generalize in an innocuous if superficially complicated way.

Proposition 4.2.2. *Fix constants $p \in [2, \infty)$ and $K \in (0, \infty)$. Let $\mathcal{V}$ be a finite-dimensional Banach space, let $\mathcal{S}$ be a block algebra and let $\mathcal{G}$ be a σ-field. Let $Y \in \mathrm{Mat}_{1 \times n}(\mathcal{S})$ and $Z \in \mathrm{Mat}_{n \times 1}(\mathcal{S})$ be random such that*

$$(\vee \|[\![Y(1,j)]\!]\|_{2p}) \vee (\vee \|[\![Z(i,1)]\!]\|_{2p}) \leq K \text{ and } \mathbb{E} Y = 0 = \mathbb{E} Z.$$

Assume also that the family $\mathcal{G} \cup \{\sigma(Y(1,i), Z(i,1))\}_{i=1}^n$ of σ-fields is independent. Then for any $\mathcal{G}$-measurable random bilinear map

$$R \in B(\mathrm{Mat}_{1 \times n}(\mathcal{S}), \mathrm{Mat}_{n \times 1}(\mathcal{S}); \mathcal{V})$$

such that $\|[\![R]\!]\|_p < \infty$ we have

$$\|[\![R(Y,Z) - \mathbb{E}(R(Y,Z)|\mathcal{G})]\!]\|_p \leq C K^2 \|[\![R]\!]\|_p \sqrt{n}$$

where the constant C depends only on p, $\mathcal{S}$ and $\mathcal{V}$.

We need two lemmas, the first of which actually proves more than we immediately need but has several further uses in the paper.

Lemma 4.2.3. *Let $\mathcal{S}$ be a block algebra of dimension s^2. (i) For $X \in \mathrm{Mat}_{k \times \ell}(\mathcal{S})$, we have*

$$\frac{1}{s} [\![X]\!]^2 \leq \sum_{i=1}^k \sum_{j=1}^\ell [\![X(i,j)]\!]^2 \leq s(k \wedge \ell) [\![X]\!]^2 .$$

(ii) For $X \in \mathrm{Mat}_{k \times \ell}(\mathcal{S})$ and $Y \in \mathrm{Mat}_{\ell \times k}(\mathcal{S})$, we have

$$[\![\mathrm{tr}_{\mathcal{S}}(XY)]\!] \leq s\ell [\![X]\!] [\![Y]\!] .$$

(iii) For $X \in \mathrm{Mat}_n(\mathcal{S})$, we have

$$\left[\!\!\left[\sum_{i,j=1}^n X(i,j)^{\otimes 2} \right]\!\!\right] \leq sn [\![X]\!]^2 .$$

Proof. Statement (i) is an assertion concerning the Hilbert-Schmidt norm which is easy to verify. Statements (ii) and (iii) follow from statement (i) via the Cauchy-Schwarz inequality. $\square$

Lemma 4.2.4. *Let $\mathcal{S}$ be a block algebra and let $\{e_i\}_{i=1}^{\ell}$ be a basis of the underlying vector space. Let $\mathcal{V}$ be a finite-dimensional Banach space and let $\{v_k\}_{k=1}^{m}$ be a basis of the underlying vector space. Fix matrices $R_{ij}^k \in \mathrm{Mat}_n(\mathbb{C})$ for $i,j = 1,\dots,\ell$ and $k = 1,\dots,m$. Define $R \in B(\mathrm{Mat}_{1\times n}(\mathcal{S}), \mathrm{Mat}_{n\times 1}(\mathcal{S}); \mathcal{V})$ by requiring that*

$$R\,(x \otimes e_i, y \otimes e_j) = \sum_k (x R_{ij}^k y) v_k$$

for $i,j = 1,\dots,\ell$, $x \in \mathrm{Mat}_{1\times n}(\mathbb{C})$ and $y \in \mathrm{Mat}_{n\times 1}(\mathbb{C})$. Then

$$\frac{1}{C} \bigvee_{i,j,k} [\![R_{ij}^k]\!] \leq [\![R]\!] \leq C \sum_{i,j,k} [\![R_{ij}^k]\!]$$

for a constant $C \geq 1$ which depends only on the data $(\mathcal{S}, \{e_i\}, \mathcal{V}, \{v_k\})$ and in particular is independent of n.

Proof. By Lemma 4.1.7 and the fact that the map

$$(A \mapsto ((x,y) \mapsto xAy)) : \mathrm{Mat}_n(\mathbb{C}) \to B(\mathrm{Mat}_{1\times n}(\mathbb{C}), \mathrm{Mat}_{n\times 1}(\mathbb{C}); \mathbb{C})$$

is an isometric isomorphism, the proof of the lemma at hand reduces to a straightforward calculation the remaining details of which we can safely omit. □

4.2.5. *Proof of Proposition 4.2.2.* After using standard properties of conditional expectation we may assume that R is deterministic. We may also assume that $\mathcal{S}$ is isomorphic to $\mathrm{Mat}_s(\mathbb{C})$ for some s and in turn Lemma 4.2.4 permits us to assume that $\mathcal{S} = \mathbb{C}$. Finally, by Lemma 4.2.3, the proposition at hand reduces to Proposition 4.2.1. □

Remark 4.2.6. In applications of Proposition 4.2.2 we will only use two special types of bilinear map R. We describe these types and estimate $[\![R]\!]$ for each. (They conform to the patterns sets by the objects Q^N_{I,J,j_1,j_2} and P^N_{I,J,j_1,j_2} defined in §8 below, respectively.) (i) In the "Q-type" first case of interest, we have $\mathcal{V} = \mathcal{S}$ and for some $A \in \mathrm{Mat}_n(\mathcal{S})$ we have $R(y,z) = yAz$, in which case $[\![R]\!] \leq [\![A]\!]$. (ii) In the "P-type" second case of interest, we have $\mathcal{V} = B(\mathcal{S})$, and for some $A \in \mathrm{Mat}_n(\mathcal{S})$ we have $R(y,z) = (B \mapsto \mathrm{tr}_{\mathcal{S}}(AzByA))$, in which case $[\![R]\!] \leq s\,[\![A]\!]^2$ for s equal to the square root of the dimension of $\mathcal{S}$ over the complex numbers by Lemma 4.2.3.

4.3. **A conditional variance bound.** We present a result which harmlessly generalizes the well-known subadditivity of variance to a situation involving vector-valued random variables and some mild dependence.

4.3.1. *Setup for the result.* Let $\mathcal{V}$ be a finite-dimensional Banach space (either real or complex scalars). Let $\{\mathcal{E}\} \cup \{\mathcal{G}(i,j)\}_{1\leq i\leq j\leq N}$ be a family of independent σ-fields and let $\mathcal{G}$ be the σ-field generated by this family. Let $Z \in \mathcal{V}$ be a $\mathcal{G}$-measurable random vector such that $\|[\![Z]\!]\|_p < \infty$ for $p \in [1,\infty)$. For $k = 1,\dots,N$, let $\widehat{\mathcal{G}}_k$ be the σ-field generated by the subfamily $\{\mathcal{E}\} \cup \{\mathcal{G}(i,j) \mid k \notin \{i,j\}\}$ and let $Z_k \in \mathcal{V}$ be a $\widehat{\mathcal{G}}_k$-measurable random vector such that $\|[\![Z_k]\!]\|_p < \infty$ for $p \in [1,\infty)$.

Proposition 4.3.2. *Notation and assumptions are as above. For every constant $p \in [1,\infty)$ we have*

$$\left\|\mathbb{E}([\![Z - \mathbb{E}(Z|\mathcal{E})]\!]^2\,|\mathcal{E})\right\|_p \leq c \sum_{k=1}^N \|[\![Z - Z_k]\!]\|_{2p}^2, \tag{52}$$

for a constant c depending only on $\mathcal{V}$ and in particular independent of p.

Proof. We may assume that $\mathcal{V}$ is a (finite-dimensional) real Hilbert space, and in this case we will prove the claim with a constant $c = 1$. After a routine application of Minkowski and Jensen inequalities, it is enough to prove

$$\mathbb{E}(\llbracket Z - \mathbb{E}(Z|\mathcal{E})\rrbracket^2 \,|\mathcal{E}) \leq \sum_{k=1}^{N} \mathbb{E}(\llbracket Z - Z_k\rrbracket^2 \,|\mathcal{E}), \tag{53}$$

almost surely. There is also no harm in assuming that $\mathcal{V} = \mathbb{R}$. For $k = 0, \ldots, N$, let $\mathcal{G}_k$ be the σ-field generated by the subfamily $\{\mathcal{E}\} \cup \{\mathcal{G}(i, j) \mid 1 \leq i \leq j \leq k\}$. In any case, by orthogonality of martingale increments, we have

$$\mathbb{E}(\llbracket Z - \mathbb{E}(Z|\mathcal{E})\rrbracket^2 \,|\mathcal{E}) = \sum_{i=1}^{N} \mathbb{E}(\llbracket \mathbb{E}(Z|\mathcal{G}_k) - \mathbb{E}(Z|\mathcal{G}_{k-1})\rrbracket^2 \,|\mathcal{E}),$$

almost surely. Furthermore we have

$$\mathbb{E}(\mathbb{E}(Z|\widehat{\mathcal{G}}_k)|\mathcal{G}_k) = \mathbb{E}(Z|\mathcal{G}_{k-1}),$$

almost surely. Finally, we have

$$\begin{aligned}\mathbb{E}(\llbracket \mathbb{E}(Z|\mathcal{G}_k) - \mathbb{E}(Z|\mathcal{G}_{k-1})\rrbracket^2 \,|\mathcal{E}) \;&=\; \mathbb{E}\left(\left\llbracket \mathbb{E}(Z - \mathbb{E}(Z|\widehat{\mathcal{G}}_k)|\mathcal{G}_k)\right\rrbracket^2 \middle|\, \mathcal{E}\right)\\ \leq\; \mathbb{E}\left(\left\llbracket Z - \mathbb{E}(Z|\widehat{\mathcal{G}}_k)\right\rrbracket^2 \middle|\, \mathcal{E}\right) \;&\leq\; \mathbb{E}(\llbracket Z - Z_k\rrbracket^2 \,|\mathcal{E}),\end{aligned}$$

almost surely, whence (53). $\square$

Definition 4.3.3. The random variable $\mathbb{E}(\llbracket Z - \mathbb{E}(Z|\mathcal{E})\rrbracket^2 \,|\mathcal{E})$ appearing on the left side of (52) will be denoted by $\mathbf{Var}_{\mathcal{V}}(Z|\mathcal{E})$ in the sequel.

4.4. **Estimates for tensor-cubic forms.** We work out a specialized estimate involving three-fold tensor products and partitions of a set of cardinality six. The combinatorial apparatus introduced here will have further uses.

4.4.1. *Set partitions and related apparatus.* A *set partition* of k is a disjoint family Π of nonempty subsets of the set $\{1, \ldots, k\}$ whose union is $\{1, \ldots, k\}$. Each member of a set partition is called a *part*. Let $\mathrm{Part}(k)$ be the family of set partitions of k. Let $\mathrm{Part}^*(2k)$ be the subset of $\mathrm{Part}(2k)$ consisting of set partitions having no singleton as a part, nor having any of the sets $\{2i-1, 2i\}$ for $i = 1, \ldots, k$ as a part. Let $\mathrm{Part}^*_2(2k) \subset \mathrm{Part}^*(2k)$ be the subfamily consisting of partitions all of whose parts have cardinality 2. For each positive integer k let S_k be the group of permutations of $\{1, \ldots, k\}$. Let $\Gamma_k \subset S_{2k}$ be the subgroup centralizing the involutive permutation $(12)\cdots(2k-1, 2k)$. Then Γ_k acts on the set $\mathrm{Part}^*(2k)$. For $\Pi_1, \Pi_2 \in \mathrm{Part}^*(2k)$ belonging to the same Γ_k-orbit we write $\Pi_1 \sim \Pi_2$.

4.4.2. *Explicit descriptions of* $\mathrm{Part}^*(4)$ *and* $\mathrm{Part}^*(6)$. To describe $\mathrm{Part}^*(4)$ we can easily enumerate it, thus:

$$\{\{1, 2, 3, 4\}\}, \{\{1, 3\}, \{2, 4\}\}, \{\{1, 4\}, \{2, 3\}\}. \tag{54}$$

It can be shown (we omit the tedious details) that for every $\Pi \in \mathrm{Part}^*(6)$ there exists exactly one set partition on the list

$$\begin{aligned}&\{\{1, 2, 3, 4, 5, 6\}\},\ \{\{1, 6\}, \{2, 3, 4, 5\}\},\ \{\{1, 3, 5\}, \{2, 4, 6\}\},\\ &\{\{1, 6\}, \{2, 3\}, \{4, 5\}\},\ \ \{\{1, 2, 3\}, \{4, 5, 6\}\}\end{aligned} \tag{55}$$

belonging to the Γ_3-orbit of Π.

4.4.3. *Sequences and associated partitions.* For any finite set I we write

$$\mathrm{Seq}(k,I)=\{\mathbf{i}:\{1,\dots,k\}\to I\}.$$

Given $\mathbf{i}\in\mathrm{Seq}(k,I)$, let $\Pi(\mathbf{i})\in\mathrm{Part}(k)$ be the set partition generated by $\mathbf{i}$, i.e., the coarsest set partition on the parts of which $\mathbf{i}$ is constant. If $I=\{1,\dots,n\}$ we write $\mathrm{Seq}(k,I)=\mathrm{Seq}(k,n)$ by abuse of notation. Sometimes we represent elements of $\mathrm{Seq}(k,I)$ as "words" $i_1\cdots i_k$ spelled with "letters" $i_1,\dots,i_k\in I$.

4.4.4. *Setup for the main result.* Let $\mathcal{S}$ be a block algebra. Let a set partition $\Pi\in\mathrm{Part}^*(6)$ and matrices $M_1,M_2,M_3\in\mathrm{Mat}_n(\mathcal{S})$ be given. Put

$$\mathfrak{M}_\Pi=\begin{cases}\left[\!\!\left[\displaystyle\sum_{\substack{\mathbf{i}=i_1\cdots i_6\\ \in\mathrm{Seq}(6,n)\\ \text{s.t. }\Pi(\mathbf{i})=\Pi}} M_1(i_1,i_2)\otimes M_2(i_3,i_4)\otimes M_3(i_5,i_6)\right]\!\!\right] & \text{if } \Pi\in\mathrm{Part}_2^*(6),\\[2ex] \displaystyle\sum_{\substack{\mathbf{i}=i_1\cdots i_6\\ \in\mathrm{Seq}(6,n)\\ \text{s.t. }\Pi(\mathbf{i})=\Pi}} [\![M_1(i_1,i_2)]\!]\,[\![M_2(i_3,i_4)]\!]\,[\![M_3(i_5,i_6)]\!] & \text{if } \Pi\not\in\mathrm{Part}_2^*(6).\end{cases}$$

Proposition 4.4.5. *Notation and assumptions are as above. For* $\Pi\in\mathrm{Part}^*(6)$, *unless* $\Pi\sim\{\{1,2,3\},\{4,5,6\}\}$, *we have* $\mathfrak{M}_\Pi\le cn\,[\![M_1]\!]\,[\![M_2]\!]\,[\![M_3]\!]$ *for a constant* c *depending only on* $\mathcal{S}$.

Proof. We may assume that $\mathcal{S}$ is isomorphic to $\mathrm{Mat}_s(\mathbb{C})$ and thus by Lemma 4.1.7 that $\mathcal{S}=\mathbb{C}$. After replacing (M_1,M_2,M_3) by $(M_{\sigma(1)}^{\mathrm{T}^{\nu_1}},M_{\sigma(2)}^{\mathrm{T}^{\nu_2}},M_{\sigma(3)}^{\mathrm{T}^{\nu_3}})$ for suitably chosen $\sigma\in S_3$ and $\nu_1,\nu_2,\nu_3\in\{0,1\}$, we may assume that Π appears on the list (55). We may also assume that each matrix M_α is either diagonal or else vanishes identically on the diagonal. Finally, we may assume that $\mathfrak{M}_\Pi>0$. Let d be the number of matrices M_α which are diagonal. Consider the following mutually exclusive and exhaustive collection of cases:

(i) $\Pi=\{\{1,6\},\{2,3\},\{4,5\}\}$ and hence $d=0$.
(ii) $\Pi=\{\{1,3,5\},\{2,4,6\}\}$ and hence $d=0$.
(iii) $\Pi=\{\{1,6\},\{2,3,4,5\}\}$ and hence $d=1$.
(iv) $\Pi=\{\{1,2,3,4,5,6\}\}$ and hence $d=3$.

In case (i) we have $\mathfrak{M}_\Pi=|\mathrm{tr}\,M_1M_2M_3|\le n\,[\![M_1M_2M_3]\!]\le n\,[\![M_1]\!]\,[\![M_2]\!]\,[\![M_3]\!]$. In case (ii) we have

$$\begin{aligned}\mathfrak{M}_\Pi&\le [\![M_1]\!]\textstyle\sum_{i,j=1}^n|M_2(i,j)M_3(i,j)|\\ &\le [\![M_1]\!]\textstyle\prod_{\alpha\in\{2,3\}}\left(\sum_{i,j=1}^n|M_\alpha(i,j)|^2\right)^{1/2}\le n\,[\![M_1]\!]\,[\![M_2]\!]\,[\![M_3]\!].\end{aligned}$$

In case (iii), similarly, we have

$$\mathfrak{M}_\Pi\le[\![M_2]\!]\textstyle\sum_{i,j=1}^n|M_1(i,j)M_3(j,i)|\le n\,[\![M_1]\!]\,[\![M_2]\!]\,[\![M_3]\!].$$

Finally, in case (iv) we have $\mathfrak{M}_\Pi\le n\,[\![M_1]\!]\,[\![M_2]\!]\,[\![M_3]\!]$ simply by counting. $\square$

5. Transpositions, SALT block designs and the secondary trick

We introduce algebraic tools for exploiting the symmetry which the random matrices (33) possess as a consequence of assumption (17) and which the operators (34) also possess by virtue of the special choice of Ξ made in §3.4. We introduce SALT block designs, we show how the self-adjoint linearization trick generates examples of such, and we use SALT block designs to put the crucial equations (29) and (30) into streamlined form. See Remark 5.4.4 below. We introduce the secondary trick which produces new SALT block designs from old and in particular produces some SALT block designs which do not come from the self-adjoint linearization trick.

5.1. **Transpositions.**

Definition 5.1.1. Let $\mathcal{A}$ be a $*$-algebra. A *transposition* $a \mapsto a^{\mathrm{T}}$ of $\mathcal{A}$ is a $\mathbb{C}$-linear map such that $(a^{\mathrm{T}})^{\mathrm{T}} = a$, $(a^*)^{\mathrm{T}} = (a^{\mathrm{T}})^*$ and $(ab)^{\mathrm{T}} = b^{\mathrm{T}}a^{\mathrm{T}}$ for all $a, b \in \mathcal{A}$. Necessarily $1_{\mathcal{A}}^{\mathrm{T}} = 1_{\mathcal{A}}$. A $*$-algebra (resp., C^*-algebra) equipped with a transposition T will be called a $(*, \mathrm{T})$-*algebra* (resp., $C^{*,\mathrm{T}}$-*algebra*).

Remark 5.1.2. Of course $\mathrm{Mat}_n(\mathbb{C})$ is a $C^{*,\mathrm{T}}$-algebra. More generally, for any Hilbert space H equipped with an orthonormal basis $\{h_i\}$, there exists a unique structure of $C^{*,\mathrm{T}}$-algebra for $B(H)$ such that $(h_i, Ah_j) = (h_j, A^{\mathrm{T}}h_i)$ for all operators $A \in B(H)$ and indices i and j. The concept of $C^{*,\mathrm{T}}$-algebra is essentially equivalent to that of a real C^*-algebra.

5.1.3. *Transpositions, tensor products and matrices.* Given $(*, \mathrm{T})$-algebras $\mathcal{A}$ and $\mathcal{B}$, we invariably equip $\mathcal{A} \otimes \mathcal{B}$ with a transposition by the rule $(a \otimes b)^{\mathrm{T}} = a^{\mathrm{T}} \otimes b^{\mathrm{T}}$, thus equipping $\mathcal{A} \otimes \mathcal{B}$ with the structure of $(*, \mathrm{T})$-algebra. Note that if $\mathcal{A}$ and $\mathcal{B}$ are $C^{*,\mathrm{T}}$-algebras at least one of which is finite-dimensional, then $\mathcal{A} \otimes \mathcal{B}$ is again a $C^{*,\mathrm{T}}$-algebra. For any $(*, \mathrm{T})$-algebra $\mathcal{A}$ and matrix $A \in \mathrm{Mat}_{k\times\ell}(\mathcal{A})$, we define $A^{\mathrm{T}} \in \mathrm{Mat}_{\ell\times k}(\mathcal{A})$ by $A^{\mathrm{T}}(i, j) = A(j, i)^{\mathrm{T}}$. Thus, in particular, $\mathrm{Mat}_n(\mathcal{A})$ is automatically a $(*, \mathrm{T})$-algebra (resp., $C^{*,\mathrm{T}}$-algebra) whenever $\mathcal{A}$ is.

5.1.4. *Transpositions on* $\mathbb{C}\langle \mathbf{X}\rangle$ *and* $B(\mathcal{H})$. We equip the noncommutative polynomial algebra $\mathbb{C}\langle \mathbf{X}\rangle$ with a transposition by the rule $\mathbf{X}_\ell^{\mathrm{T}} = (-1)^\ell \mathbf{X}_\ell$ for every ℓ. Note that the C^*-algebra $B(\mathcal{H})$ is canonically equipped with a transposition because Boltzmann-Fock space $\mathcal{H}$ is canonically equipped with an orthonormal basis.

Remark 5.1.5. We claim that the evaluation maps

$$\left(f \mapsto f\left(\frac{\Xi^N}{\sqrt{N}}\right)\right) : \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X}\rangle) \to \mathrm{Mat}_{nN}(\mathbb{C}),$$
$$(f \mapsto f(\Xi)) : \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X}\rangle) \to \mathrm{Mat}_n(B(\mathcal{H}))$$

figuring in Theorems 1.2.4, 1.2.5 and 2.4.6 are $(*, \mathrm{T})$-algebra homomorphisms. In any case, it is clear that each homomorphism is a $*$-algebra homomorphism, so we have only to verify that each map commutes with the transposition. The former commutes with T by assumption (17) which (recall) says that $(\Xi_\ell^N)^{\mathrm{T}} = (-1)^\ell \Xi_\ell^N$ for all ℓ. The latter commutes with T because (recall from §3.4) by definition $\Xi_\ell = \mathrm{i}^\ell \Sigma_\ell + \mathrm{i}^{-\ell}\Sigma_\ell^*$, clearly $\Sigma_\ell^{\mathrm{T}} = \Sigma_\ell^*$, and hence $\Xi_\ell^{\mathrm{T}} = (-1)^\ell \Xi_\ell$ for all ℓ. The claim is proved.

Lemma 5.1.6. *If x is an element of a $C^{*,\mathrm{T}}$-algebra $\mathcal{A}$, then $(x^{-1})^{\mathrm{T}} = (x^{\mathrm{T}})^{-1}$, $x \in \mathcal{A}_{\mathrm{sa}} \Rightarrow x^{\mathrm{T}} \in \mathcal{A}_{\mathrm{sa}}$, $\mathrm{Spec}(x) = \mathrm{Spec}(x^{\mathrm{T}})$, $x \geq 0 \Rightarrow x^{\mathrm{T}} \geq 0$ and $[\![x^{\mathrm{T}}]\!] = [\![x]\!]$.*

Proof. The first two claims are obvious. The third claim follows from the first. The second and third claims imply the fourth. The fifth holds for self-adjoint x by Proposition 3.2.4 along with the second and third claims. The fifth claim holds in general because $\left[\!\left[x^{\mathrm{T}}\right]\!\right]^2 = \left[\!\left[(x^{\mathrm{T}})^* x^{\mathrm{T}}\right]\!\right] = \left[\!\left[(xx^*)^{\mathrm{T}}\right]\!\right] = [\![xx^*]\!] = [\![x]\!]^2$. $\square$

Definition 5.1.7. Given a $C^{*,\mathrm{T}}$-algebra $\mathcal{A}$ and a state $\phi \in \mathcal{A}^\star$, we say that ϕ is T-*stable* if $\phi(A^{\mathrm{T}}) = \phi(A)$ for all $A \in \mathcal{A}$. A pair $(\mathcal{A}, \phi)$ consisting of a $C^{*,\mathrm{T}}$-algebra and a T-stable state ϕ will be called a $C^{*,\mathrm{T}}$-*probability space*.

Remark 5.1.8. It is easy to see that both $(\mathrm{Mat}_N(\mathbb{C}), \frac{1}{N}\mathrm{tr})$ and $(B(\mathcal{H}), \phi^{\mathrm{BF}})$ are in fact $C^{*,\mathrm{T}}$-probability spaces.

5.2. **Block algebras (enhanced version).** We re-introduce the notion of block algebra, this time with structure enriched by a transposition. We also introduce the notion of $\mathcal{S}$-linear form in terms of which we will handle much bookkeeping below.

Definition 5.2.1. A *block algebra* is a $C^{*,\mathrm{T}}$-algebra isomorphic to the matrix algebra $\mathrm{Mat}_s(\mathbb{C})$ for some integer $s > 0$. A basis $\{e_{ij}\}_{i,j=1}^s$ for $\mathcal{S}$ such that $e_{ij}e_{i'j'} = \delta_{ji'}e_{ij'}$ and $e_{ij}^* = e_{ji} = e_{ij}^{\mathrm{T}}$ will be called *standard*.

Remark 5.2.2. A choice of standard basis of a block algebra is the same thing as a choice of a $C^{*,\mathrm{T}}$-algebra isomorphism with $\mathrm{Mat}_s(\mathbb{C})$.

Remark 5.2.3. The tensor product of block algebras is again a block algebra. Furthermore, for every block algebra $\mathcal{S}$, the tensor product algebra $\mathbb{C}\langle\mathbf{X}\rangle \otimes \mathcal{S}$ (resp., $B(\mathcal{H}) \otimes \mathcal{S}$) is a $(*, \mathrm{T})$-algebra (resp., $C^{*,\mathrm{T}}$-algebra).

Definition 5.2.4. Let $\mathcal{S}$ be any block algebra. An $\mathcal{S}$-*linear form* L is an element of the tensor product algebra $\mathbb{C}\langle\mathbf{X}\rangle \otimes \mathcal{S}$ of the form $L = \sum_{\ell=1}^\infty \mathbf{X}_\ell \otimes a_\ell$ for some elements $a_\ell \in \mathcal{S}$ vanishing for $\ell \gg 0$. We refer to the sum $\sum_\ell \mathbf{X}_\ell \otimes a_\ell$ as the *Hamel expansion* of L and to the elements $a_\ell \in \mathcal{S}$ as the *Hamel coefficients* of L. Given a sequence $\xi = \{\xi_\ell\}_{\ell=1}^\infty \in \mathcal{A}^\infty$ in an algebra $\mathcal{A}$, we define $L(\xi) = \sum_\ell \xi_\ell \otimes a_\ell \in \mathcal{A} \otimes \mathcal{S}$, calling this the *evaluation of* L *at* ξ. It is especially important to notice that if $\mathcal{A} = \mathrm{Mat}_N(\mathbb{C})$, then $L(\xi) \in \mathrm{Mat}_N(\mathbb{C}) \otimes \mathcal{S} = \mathrm{Mat}_N(\mathcal{S})$. This is the reason for putting the tensor factors in $\mathbb{C}\langle\mathbf{X}\rangle \otimes \mathcal{S}$ in the "wrong" order.

Definition 5.2.5. Let $\mathcal{S}$ be a block algebra and let L be an $\mathcal{S}$-linear form with Hamel expansion $L = \sum \mathbf{X}_\ell \otimes a_\ell$. We define $\Phi_L \in B(\mathcal{S})$ by the formula $\Phi_L(\zeta) = \sum a_\ell \zeta a_\ell$ for $\zeta \in \mathcal{S}$ and we define $\Psi_L = \sum (-1)^\ell a_\ell^{\otimes 2} \in \mathcal{S}^{\otimes 2}$. We call Φ_L the *covariance map* attached to L. We call Ψ_L the *covariance tensor* attached to L.

Definition 5.2.6. Each block algebra $\mathcal{S}$ is equipped with a unique state $\tau_{\mathcal{S}}$ satisfying $\tau_{\mathcal{S}}(e_{ij}) = (\dim \mathcal{S})^{-1/2}\delta_{ij}$ for any standard basis $\{e_{ij}\}$. Necessarily $\tau_{\mathcal{S}}$ is T-stable. More generally, for each projection $e \in \mathcal{S}$, there exists a unique state $\tau_{\mathcal{S},e} \in \mathcal{S}^\star$ such that $\tau_{\mathcal{S},e}|_{e\mathcal{S}e} = \tau_{e\mathcal{S}e}$.

Remark 5.2.7. If $\mathcal{S} = \mathrm{Mat}_s(\mathbb{C})$ and $e = \begin{bmatrix} \mathbf{I}_n & 0 \\ 0 & 0 \end{bmatrix}$, then $\tau_{S,e}(A) = \frac{1}{n}\sum_{i=1}^n A(i,i)$.

5.2.8. *The bullet map.* Given a block algebra $\mathcal{S}$, we define a linear isomorphism $(A \mapsto A^\bullet) : \mathcal{S}^{\otimes 2} \to B(\mathcal{S})$ by the formula $(x \otimes y)^\bullet = (z \mapsto xzy)$. That the bullet map is indeed a linear isomorphism one can check by calculating with a standard basis. The bullet map in general neither preserves norms nor algebra structure.

5.2.9. *The half-transpose map.* Given a block algebra $\mathcal{S}$, we define a linear isomorphism $(A \mapsto A^{1\otimes \mathrm{T}}) \in B(\mathcal{S}^{\otimes 2})$ by the formula $(x \otimes y)^{1\otimes \mathrm{T}} = x \otimes y^{\mathrm{T}}$. The half-transpose map in general neither preserves norms nor algebra structure.

Remark 5.2.10. Strangely enough, the composite map

$$((x \otimes y) \mapsto ((x \otimes y)^{1\otimes \mathrm{T}})^{\bullet}) : \mathcal{S}^{\otimes 2} \to B(\mathcal{S})$$

is an isomorphism of algebras, as one verifies by calculating with a standard basis. (But this map still does not in general preserve norms.)

5.3. $\mathcal{S}$-(bi)linear constructions.

5.3.1. *$\mathcal{S}$-linear extension of states.* Given any C^*-probability space $(\mathcal{A}, \phi)$ and block algebra $\mathcal{S}$, we define the *$\mathcal{S}$-linear extension* $\phi_{\mathcal{S}} : \mathcal{A} \otimes \mathcal{S} \to \mathcal{S}$ of ϕ by the formula

$$\phi_{\mathcal{S}}(x \otimes y) = \phi(x)y.$$

Note that since ϕ commutes with the involution, the same is true for $\phi_{\mathcal{S}}$, i.e.,

$$\phi_{\mathcal{S}}(A^*) = \phi_{\mathcal{S}}(A)^* \tag{56}$$

for $A \in \mathcal{A} \otimes \mathcal{S}$. Suppose now that $(\mathcal{A}, \phi)$ is a $C^{*,\mathrm{T}}$-probability space. Note that since ϕ is T-stable, $\phi_{\mathcal{S}}$ commutes with T, i.e.,

$$\phi_{\mathcal{S}}(A^{\mathrm{T}}) = \phi_{\mathcal{S}}(A)^{\mathrm{T}} \tag{57}$$

for $A \in \mathcal{A} \otimes \mathcal{S}$.

Remark 5.3.2. Objects like $\phi_{\mathcal{S}}$ are the stock-in-trade of operator-valued free probability theory. See [13] for a useful introduction to this point of view and see [18] for in-depth treatment.

Remark 5.3.3. Consider the case $(\mathcal{A}, \phi) = (\mathrm{Mat}_N(\mathbb{C}), \frac{1}{N}\mathrm{tr})$. We have

$$\phi_{\mathcal{S}} = \left(A \mapsto \frac{1}{N}\mathrm{tr}_{\mathcal{S}} A = \sum_{i=1}^{N} A(i,i)\right) : \mathrm{Mat}_N(\mathcal{S}) \to \mathcal{S}.$$

Thus the *ad hoc* construction $\mathrm{tr}_{\mathcal{S}}$ fits into a more general conceptual framework.

Remark 5.3.4. The $\mathcal{S}$-linear extension $\phi^{\mathrm{BF}}_{\mathcal{S}}$ of the state ϕ^{BF} with which $B(\mathcal{H})$ is canonically equipped satisfies

$$(p_{\mathcal{H}} \otimes 1_{\mathcal{S}})A(p_{\mathcal{H}} \otimes 1_{\mathcal{S}}) = p_{\mathcal{H}} \otimes \phi^{\mathrm{BF}}_{\mathcal{S}}(A), \text{ hence } \left[\!\left[\phi^{\mathrm{BF}}_{\mathcal{S}}(A)\right]\!\right] \leq [\![A]\!] \tag{58}$$

for all $A \in B(\mathcal{H}) \otimes \mathcal{S}$ and hence $\left[\!\left[\phi^{\mathrm{BF}}_{\mathcal{S}}\right]\!\right] = 1$. In fact, in full generality, we have $[\![\phi_{\mathcal{S}}]\!] = 1$ by a similar argument using the GNS construction, which we omit.

Definition 5.3.5. For any $(*, \mathrm{T})$-algebra $\mathcal{A}$, let $\mathcal{A}^{\infty}_{\mathrm{alt}}$ denote the space of sequences $\xi = \{\xi_\ell\}_{\ell=1}^{\infty}$ in $\mathcal{A}$ such that $\xi_\ell^{\mathrm{T}} = (-1)^\ell \xi_\ell$ for all ℓ. Also put $\mathcal{A}^{\infty}_{\mathrm{salt}} = \mathcal{A}^{\infty}_{\mathrm{sa}} \cap \mathcal{A}^{\infty}_{\mathrm{alt}}$.

Remark 5.3.6. Let $\mathcal{A}$ be a $(*, \mathrm{T})$-algebra, $\xi = \{\xi_\ell\}_{\ell=1}^{\infty} \in \mathcal{A}^{\infty}_{\mathrm{salt}}$ a sequence and L an $\mathcal{S}$-linear form. Then we have $L^{\mathrm{T}}(\xi) = L(\xi)^{\mathrm{T}}$ and $L^*(\xi) = L(\xi)^*$. In particular, this observation applies to the sequences $\Xi^N \in \mathrm{Mat}_N(\mathbb{C})^{\infty}_{\mathrm{salt}}$ and $\Xi \in B(\mathcal{H})^{\infty}_{\mathrm{salt}}$ figuring in Theorem 2.4.6.

5.3.7. *$\mathcal{S}$-bilinear extension of states.* Let $\mathcal{S}$ be a block algebra and let $(\mathcal{A},\phi)$ be a C^*-probability space. We define the *$\mathcal{S}$-bilinear extension*

$$\phi_{\mathcal{S},\mathcal{S}}:\mathcal{A}\otimes\mathcal{S}\times\mathcal{A}\otimes\mathcal{S}\to\mathcal{S}^{\otimes 2}$$

of ϕ by the formula

$$\phi_{\mathcal{S},\mathcal{S}}(x_1\otimes y_1,x_2\otimes y_2)=\phi(x_1x_2)y_1\otimes y_2.$$

Remark 5.3.8. Let $\mathcal{S}$ be a block algebra and consider the C^*-probability space $(\mathrm{Mat}_N(\mathbb{C}),\frac{1}{N}\mathrm{tr})$. For $R_1,R_2\in\mathrm{Mat}_N(\mathcal{S})$ we have

$$\phi_{\mathcal{S},\mathcal{S}}(R_1,R_2)=\frac{1}{N}\sum_{i,j=1}^N R_1(i,j)\otimes R_2(j,i)\in\mathcal{S}^{\otimes 2}.$$

Remark 5.3.9. Consider the C^*-algebra embeddings

$$\left.\begin{array}{l}\iota^{(1)}=(x\otimes y\mapsto x\otimes y\otimes 1_{\mathcal{S}})\\ \iota^{(2)}=(x\otimes y\mapsto x\otimes 1_{\mathcal{S}}\otimes y)\end{array}\right\}:\mathcal{A}\otimes\mathcal{S}\to\mathcal{A}\otimes\mathcal{S}^{\otimes 2}.$$

One has

$$\phi_{\mathcal{S},\mathcal{S}}(A,B)=\phi_{\mathcal{S}^{\otimes 2}}(\iota^{(1)}(A)\iota^{(2)}(B))\tag{59}$$

and thus $[\![\phi_{\mathcal{S},\mathcal{S}}]\!]=1$ since $[\![\phi_{\mathcal{S}^{\otimes 2}}]\!]=1$. In a similar vein, we have the formula

$$\phi_{\mathcal{S},\mathcal{S}}(A,B)^\bullet(\zeta)=\phi_{\mathcal{S}}(A(1_{\mathcal{A}}\otimes\zeta)B)\tag{60}$$

which we will use below in §6 to study the secondary Schwinger-Dyson equation.

Remark 5.3.10. Consider the $C^{*,\mathrm{T}}$-probability space $(\mathcal{A},\phi)=(\mathrm{Mat}_N(\mathbb{C}),\frac{1}{N}\mathrm{tr})$. Let $\mathcal{S}$ be any block algebra. Let $R\in\mathrm{Mat}_N(\mathcal{S})$ be any matrix. We have

$$\phi_{\mathcal{S},\mathcal{S}}(R,R)^\bullet = \left(\zeta\mapsto\frac{1}{N}\sum_{i,j=1}^N R(i,j)\zeta R(j,i)\right)\in B(\mathcal{S}),\tag{61}$$

$$\phi_{\mathcal{S},\mathcal{S}}(R,R^{\mathrm{T}})^{1\otimes\mathrm{T}} = \frac{1}{N}\sum_{i,j=1}^N R(i,j)^{\otimes 2}\in\mathcal{S}^{\otimes 2}.\tag{62}$$

We will study the expressions of the form on the left in §6 below. We will study expressions of the form on the right in §8 and §9 below.

5.4. **An upgrade of the self-adjoint linearization trick.**

Definition 5.4.1. A *SALT block design* is a quadruple $(\mathcal{S},L,\Theta,e)$ consisting of

- a block algebra $\mathcal{S}$,
- a self-adjoint $\mathcal{S}$-linear form L,
- an element $\Theta\in\mathcal{S}$ (perhaps not self-adjoint), and
- a projection $e\in\mathcal{S}$

such that for every $C^{*,\mathrm{T}}$-algebra $\mathcal{A}$, sequence $\xi\in\mathcal{A}^\infty_{\mathrm{salt}}$, point $z\in\mathfrak{h}$ and parameter $t\geq 0$ we have

$$L(\xi)-1_{\mathcal{A}}\otimes(\Theta+ze+\mathrm{i}t1_{\mathcal{S}})\in(\mathcal{A}\otimes\mathcal{S})^\times \text{ and}\tag{63}$$

$$\left[\!\!\left[\left(L(\xi)-1_{\mathcal{A}}\otimes(\Theta+ze+\mathrm{i}t1_{\mathcal{S}})\right)^{-1}\right]\!\!\right]\leq\frac{c_0}{2}(1+[\![L(\xi)]\!])^{c_1}(1+1/\Im z)^{c_2}\tag{64}$$

for some constants $c_0,c_1,c_2\geq 1$ depending only on $(\mathcal{S},L,\Theta,e)$ and thus independent of $\mathcal{A}$, ξ, z and t. We declare any finite constant $\mathfrak{T}\geq[\![\Im\Theta]\!]+4(1+[\![\Phi_L]\!])$ to be a *cutoff* for the design, where $\Phi_L\in B(\mathcal{S})$ is as in Definition 5.2.5.

Proposition 5.4.2. *Fix $f \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X}\rangle)_{\mathrm{sa}}$ arbitrarily. Either let $\tilde{f} \in \mathrm{Mat}_s(\mathbb{C}\langle \mathbf{X}\rangle)_{\mathrm{sa}}$ be a self-adjoint linearization of f as provided by Proposition 2.6.2 or else let $\tilde{f} = f$ if all entries of f are already of degree ≤ 1 in the variables $\mathbf{X}_\ell$. Write*

$$\tilde{f} = a_0 \otimes 1_{\mathbb{C}\langle \mathbf{X}\rangle} + \sum_{\ell=1}^{m} a_\ell \otimes \mathbf{X}_\ell$$

for $m \gg 0$ and $a_0, \ldots, a_m \in \mathrm{Mat}_s(\mathbb{C})_{\mathrm{sa}}$. Let

$$\mathcal{S} = \mathrm{Mat}_s(\mathbb{C}), \quad L = \sum_{\ell=1}^{m} \mathbf{X}_\ell \otimes a_\ell, \quad \Theta = -a_0, \quad e = \left[\begin{array}{cc} \mathbf{I}_n & 0 \\ 0 & 0 \end{array}\right].$$

Then $(\mathcal{S}, L, \Theta, e)$ is a SALT block design for which $\Theta = \Theta^*$ and $c_2 = 1$.

Proof. The case $s = n$ is easy to check using Lemma 3.2.6. We leave the details to the reader. We assume $s > n$ for the rest of the proof. Then property (63) holds by Proposition 2.6.3 and Lemma 3.2.6. The latter lemma, the inversion formula (28), Lemma 4.1.7 and the definitions yield an estimate

$$\left[\!\!\left[(L(\xi) - 1_{\mathcal{A}} \otimes (\Theta + ze))^{-1}\right]\!\!\right] \leq \frac{c_0}{4}(1 + [\![L(\xi)]\!])^{c_1}(1 + 1/\Im z),$$

for $\mathcal{A}$, ξ, and z as in (64). Estimate (64) for general $t \geq 0$ (but still with $c_2 = 1$) then follows via Lemma 3.1.1 if $\frac{c_0}{2}(1 + [\![L(\xi)]\!])^{c_1}(1 + 1/\Im z)t \leq 1$ and otherwise follows via Lemma 3.2.6. $\square$

Definition 5.4.3. For each $f \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X}\rangle)_{\mathrm{sa}}$, any SALT block design $(\mathcal{S}, L, \Theta, e)$ arising via Proposition 5.4.2 will be called a *self-adjoint linearization* of f. (Self-adjoint linearizations in the relatively naive sense defined immediately after Proposition 2.6.2 will hereafter no longer be used.)

Remark 5.4.4. (This is a continuation of Remark 2.6.4.) In the sequel we will prefer to write formulas (29) and (30) above in the following form. Fix $f \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X}\rangle)_{\mathrm{sa}}$ and let $(\mathcal{S}, L, \Theta, e)$ be any self-adjoint linearization of f. Then, with $\tau_{\mathcal{S},e}$ as in Definition 5.2.6, we have

$$S_{\mu_f^N}(z) = \tau_{\mathcal{S},e}\left(\frac{1}{N}\mathrm{tr}_{\mathcal{S}}\left(\left(L\left(\frac{\Xi^N}{\sqrt{N}}\right) - \mathbf{I}_N \otimes (\Theta + ze)\right)^{-1}\right)\right), \tag{65}$$

$$S_{\mu_f}(z) = \tau_{\mathcal{S},e} \circ \phi_{\mathcal{S}}^{\mathrm{BF}}\left(\left(L(\Xi) - 1_{B(\mathcal{H})} \otimes (\Theta + ze)\right)^{-1}\right) \tag{66}$$

for $z \in \mathfrak{h}$ (resp., for $z \in \mathbb{C}\backslash \mathrm{supp}\, \mu_f$). Furthermore, for distinct indices $i, j = 1, \ldots, N$ and arbitrary $A \in \mathcal{S}$, with the covariance map $\Phi_L \in B(\mathcal{S})$ as in Definition 5.2.5, we have

$$\begin{aligned} \Phi_L(A) &= \mathbb{E}\left(\left(L\left(\Xi^N\right)(i,j)\right) A \left(L\left(\Xi^N\right)(j,i)\right)\right) \\ &= \phi_{\mathcal{S}}(L(\Xi)(1_{B(\mathcal{H})} \otimes A)L(\Xi)). \end{aligned} \tag{67}$$

(Formula (67) is actually valid for any self-adjoint $\mathcal{S}$-linear form L whether or not it is part of a self-adjoint linearization.) One verifies the first equality of (67) by applying assumptions (17), (18), (19) and (20) which fix the covariance structure of the sequence Ξ^N. One verifies the second equality of (67) by applying the relations (39) which analogously fix the (noncommutative) covariance structure of the sequence Ξ. We also have for distinct indices $i, j = 1, \ldots, N$ the formula

$$\Psi_L = \mathbb{E}\left((L(\Xi^N)(i,j))^{\otimes 2}\right) = \phi_{\mathcal{S},\mathcal{S}}(L(\Xi), L(\Xi)) \tag{68}$$

which is proved more or less the same way as formula (67). (Formula (68) is actually valid for any $\mathcal{S}$-linear form L.)

Remark 5.4.5. We return to the setting of Definition 5.4.1. We provide some amplification of relations (63) and (64). Firstly, we have for $\zeta \in \mathcal{S}$ that

$$\begin{array}{rl} & 2\left[\!\!\left[(L(\xi) - 1_{\mathcal{A}} \otimes (\Theta + ze + \mathrm{i}t1_{\mathcal{S}}))^{-1}\right]\!\!\right] [\![\zeta]\!] \leq 1 \\ \Rightarrow & L(\xi) - 1_{\mathcal{A}} \otimes (\Theta + ze + \mathrm{i}t1_{\mathcal{S}} + \zeta) \in (\mathcal{A} \otimes \mathcal{S})^{\times} \text{ and} \\ & \left[\!\!\left[(L(\xi) - 1_{\mathcal{A}} \otimes (\Theta + ze + \mathrm{i}t1_{\mathcal{S}} + \zeta))^{-1}\right]\!\!\right] \\ \leq & 2\left[\!\!\left[(L(\xi) - 1_{\mathcal{A}} \otimes (\Theta + ze + \mathrm{i}t1_{\mathcal{S}}))^{-1}\right]\!\!\right] \end{array} \tag{69}$$

via Lemma 3.1.1. Secondly, we have

$$\left[\!\!\left[(L(\xi) - 1_{\mathcal{A}} \otimes (\Theta + ze + \mathrm{i}t1_{\mathcal{S}}))^{-1}\right]\!\!\right] \leq \frac{1}{t - [\![\Im\Theta]\!]} \text{ for } t > [\![\Im\Theta]\!] \tag{70}$$

via Lemma 3.2.6. The bound latter helps to explain the meaning of the cutoff $\mathfrak{T}$.

Remark 5.4.6. Let $(\mathcal{S}, L, \Theta, e)$ be a SALT block design and let c_0, c_1, c_2 and $\mathfrak{T}$ be the constants from Definition 5.4.1. Let $Y \in \mathrm{Mat}_N(\mathcal{S})_{\mathrm{sa}}$ be of the form $Y = L(\eta)$ for some $\eta \in \mathrm{Mat}_N(\mathbb{C})^{\infty}_{\mathrm{salt}}$. Then for every $z \in \mathfrak{h}$ and $t \in [0, \infty)$ we have

$$\begin{array}{rl} & Y - \mathbf{I}_N \otimes (\Theta + ze + \mathrm{i}t1_{\mathcal{S}}) \in \mathrm{GL}_N(\mathcal{S}) \text{ and} \\ & \left[\!\!\left[(Y - \mathbf{I}_N \otimes (\Theta + ze + \mathrm{i}t1_{\mathcal{S}}))^{-1}\right]\!\!\right] \\ \leq & \left\{ \begin{array}{ll} c_0(1 + [\![Y]\!])^{c_1}(1 + 1/\Im z)^{c_2} & \text{in general and} \\ \frac{1}{2} & \text{for } t \geq \mathfrak{T} \end{array} \right. \end{array} \tag{71}$$

by definition of a SALT block design along with Remark 5.4.5. By the resolvent identity (35), the following important (if trivial) observation also holds:

$$\begin{array}{l} \text{For each fixed } z \in \mathfrak{h}, \text{ the Lipschitz constant of the map} \\ \left(t \mapsto (Y - \mathbf{I}_N \otimes (\Theta + ze + \mathrm{i}t1_{\mathcal{S}}))^{-1}\right) : [0, \infty) \to \mathrm{Mat}_N(\mathcal{S}) \\ \text{does not exceed } c_0^2(1 + [\![Y]\!])^{2c_1}(1 + 1/\Im z)^{2c_2}. \end{array} \tag{72}$$

5.5. **The secondary trick.** We first present the underline construction and then the secondary trick itself. As a byproduct we will find natural examples of SALT block designs beyond those produced by Proposition 5.4.2. See Remark 5.5.5 below.

5.5.1. *The underline construction for a block algebra.* Let $\mathcal{M}_3$ denote a block algebra equipped with a standard basis $\{e_{ij}\}_{i,j=1}^3$. Let $\mathcal{S}$ be any block algebra. We define

$$\underline{\mathcal{S}} = \mathcal{S}^{\otimes 2} \otimes \mathcal{M}_3, \;\; \Diamond_{\mathcal{S}} = 1_{\mathcal{S}}^{\otimes 2} \otimes (e_{12} + e_{13}) \in \underline{\mathcal{S}}.$$

Furthermore, given $\Lambda \in \mathcal{S}$ we define

$$\underline{\Lambda}_1 = \Lambda \otimes 1_{\mathcal{S}} \in \mathcal{S}^{\otimes 2}, \;\; \underline{\Lambda}_2 = 1_{\mathcal{S}} \otimes \Lambda \in \mathcal{S}^{\otimes 2},$$

$$\underline{\Lambda} = \underline{\Lambda}_1 \otimes e_{11} + \underline{\Lambda}_2 \otimes e_{22} + \underline{\Lambda}_2^{\mathrm{T}} \otimes e_{33} \in \underline{\mathcal{S}}.$$

We also define linear maps

$$\partial_1 \in B(\underline{\mathcal{S}}, B(\mathcal{S})) \text{ and } \partial_2 \in B(\underline{\mathcal{S}}, \mathcal{S}^{\otimes 2})$$

by the formulas

$$\partial_1(A \otimes e_{ij}) = A^{\bullet}\delta_{1i}\delta_{2j}, \;\; \partial_2(A \otimes e_{ij}) = A^{1\otimes\mathrm{T}}\delta_{1i}\delta_{3j}$$

for $A \in \mathcal{S}^{\otimes 2}$ and $i, j = 1, \ldots, 3$.

5.5.2. *The underline construction for $\mathcal{S}$-linear forms.* Let $\mathcal{S}$ be a block algebra and let $\underline{\mathcal{S}}$ be the corresponding "underlined" block algebra as defined in the preceding paragraph. Given an $\mathcal{S}$-linear form L with Hamel expansion $L = \sum \mathbf{X}_\ell \otimes a_\ell$, we define $\mathcal{S}^{\otimes 2}$-linear forms $\underline{L}_1$ and $\underline{L}_2$ by the formulas

$$\underline{L}_1 = \sum_\ell \mathbf{X}_\ell \otimes a_\ell \otimes 1_{\mathcal{S}}, \;\; \underline{L}_2 = \sum_\ell \mathbf{X}_\ell \otimes 1_{\mathcal{S}} \otimes a_\ell$$

and in turn we define an $\underline{\mathcal{S}}$-linear form $\underline{L}$ by the formula

$$\underline{L} = \underline{L}_1 \otimes e_{11} + \underline{L}_2 \otimes e_{22} + \underline{L}_2^{\mathrm{T}} \otimes e_{33}.$$

5.5.3. *The trick itself.* Let $\mathcal{S}$ be a block algebra, let $\Lambda \in \mathcal{S}$ be an element and let L be an $\mathcal{S}$-linear form. Let $\underline{\mathcal{S}}$, $\underline{\Lambda}_1$, $\underline{\Lambda}_2$, $\underline{\Lambda}$, $\underline{L}_1$, $\underline{L}_2$ and $\underline{L}$ be as defined in the preceding two paragraphs. Let $(\mathcal{A}, \phi)$ be a $C^{*,\mathrm{T}}$-probability space and fix a sequence $\xi \in \mathcal{A}_{\mathrm{salt}}^\infty$. For the rest of this paragraph we abuse notation by writing $x = 1_{\mathcal{A}} \otimes x$ for $x \in \mathcal{S}$, $x \in \mathcal{S}^{\otimes 2}$ or $x \in \underline{\mathcal{S}}$. We assume that $L(\xi) - \Lambda \in (\mathcal{A} \otimes \mathcal{S})^\times$, in which case $\underline{L}(\xi) - \underline{\Lambda} \in (\mathcal{A} \otimes \underline{\mathcal{S}})^\times$ via Lemma 5.1.6. In turn one can verify that

$$\underline{L}(\xi) - \underline{\Lambda} - \lozenge_{\mathcal{S}} \in (\mathcal{A} \otimes \underline{\mathcal{S}})^\times$$

and more precisely

$$\begin{array}{rl} & (\underline{L}(\xi) - \underline{\Lambda} - \lozenge_{\mathcal{S}})^{-1} \\ = & (\underline{L}_1(\xi) - \underline{\Lambda}_1)^{-1} \otimes e_{11} + (\underline{L}_2(\xi) - \underline{\Lambda}_2)^{-1} \otimes e_{22} + ((\underline{L}_2(\xi) - \underline{\Lambda}_2)^{-1})^{\mathrm{T}} \otimes e_{33} \\ & +((\underline{L}_1(\xi) - \underline{\Lambda}_1)^{-1}(\underline{L}_2(\xi) - \underline{\Lambda}_2)^{-1}) \otimes e_{12} \\ & +((\underline{L}_1(\xi) - \underline{\Lambda}_1)^{-1}((\underline{L}_2(\xi) - \underline{\Lambda}_2)^{-1})^{\mathrm{T}}) \otimes e_{13}, \end{array}$$

by direct calculation. It follows that

$$[\![L(\xi)]\!] = [\![\underline{L}(\xi)]\!], \;\; \left[\!\left[(\underline{L}(\xi) - \underline{\Lambda} - \lozenge_{\mathcal{S}})^{-1}\right]\!\right] \;\leq\; 3(1 \vee \left[\!\left[(L(\xi) - \Lambda)^{-1}\right]\!\right])^2, \tag{73}$$

$$\partial_1 \circ \phi_{\underline{\mathcal{S}}}((\underline{L}(\xi) - \underline{\Lambda} - \lozenge_{\mathcal{S}})^{-1}) \;=\; \phi_{\mathcal{S},\mathcal{S}}((L(\xi) - \Lambda)^{-1}, (L(\xi) - \Lambda)^{-1})^{\bullet}, \tag{74}$$

$$\partial_2 \circ \phi_{\underline{\mathcal{S}}}((\underline{L}(\xi) - \underline{\Lambda} - \lozenge_{\mathcal{S}})^{-1}) \;=\; \phi_{\mathcal{S},\mathcal{S}}((L(\xi) - \Lambda)^{-1}, ((L(\xi) - \Lambda)^{-1})^{\mathrm{T}})^{1\otimes\mathrm{T}}, \tag{75}$$

where to get the last two identities we use the trivial formula (59).

Lemma 5.5.4. *For any SALT block design $(\mathcal{S}, L, \Theta, e)$, again $(\underline{\mathcal{S}}, \underline{L}, \underline{\Theta} + \lozenge_{\mathcal{S}}, \underline{e})$ is a SALT block design. More precisely, if c_0, c_1 and c_2 are constants rendering the estimate* (64) *valid for $(\mathcal{S}, L, \Theta, e)$, then one can take the corresponding constants $\underline{c}_0$, $\underline{c}_1$ and $\underline{c}_2$ for $(\underline{\mathcal{S}}, \underline{L}, \underline{\Theta} + \lozenge_{\mathcal{S}}, \underline{e})$ to be $\underline{c}_0 = 3c_0^2$, $\underline{c}_1 = 2c_1$ and $\underline{c}_2 = 2c_2$.*

Proof. One can read off the necessary estimates from (73). □

Remark 5.5.5. Note that $\underline{\Theta} + \lozenge_{\mathcal{S}} \in \underline{\mathcal{S}}$ is in general not self-adjoint and (more significantly) $\underline{c}_2 = 2c_2$. Thus the extra generality in Definition 5.4.1 not used by Proposition 5.4.2 is forced on us in order to make Lemma 5.5.4 hold.

6. Construction of solutions of the Schwinger-Dyson equation

We construct solutions of the Schwinger-Dyson equation by using the Boltzmann-Fock apparatus reviewed in §3 above along with the $\mathcal{S}$-linear machinery introduced in §5 above. See Proposition 6.2.2 below. Following [9], we then express the Stieltjes transform $S_{\mu_f}(z)$ figuring in Theorem 2.4.6 in terms of one of the solutions so constructed. See Remark 6.2.4 below. We also construct solutions of a secondary version of the Schwinger-Dyson equation by using the $\mathcal{S}$-bilinear machinery of §5. See Proposition 6.4.2 below. We then define our candidate for the correction $\mathrm{bias}_f^N(z)$ figuring in Theorem 2.4.6. See Remark 6.5.5 below.

6.1. **The Schwinger-Dyson equation and its differentiated form.**

Definition 6.1.1. Let $\mathcal{S}$ be a block algebra. Let $\mathcal{D} \subset \mathcal{S}$ be a (nonempty) open subset. Let $\Phi \in B(\mathcal{S})$ be a linear map. We say that an analytic function $G : \mathcal{D} \to \mathcal{S}$ satisfies the *Schwinger-Dyson (SD) equation* with *covariance map* Φ if

$$1_{\mathcal{S}} + (\Lambda + \Phi(G(\Lambda)))G(\Lambda) = 0$$

for all $\Lambda \in \mathcal{D}$. Necessarily one has $G(\Lambda) \in \mathcal{S}^\times$ for all $\Lambda \in \mathcal{D}$.

See [21], [18] or [1] for background.

6.1.2. *Notation for derivatives.* Given an analytic function $G : \mathcal{D} \to \mathcal{S}$ defined on an open subset $\mathcal{D}$ of a block algebra $\mathcal{S}$ and $\Lambda \in \mathcal{D}$, we define

$$\mathbf{D}[G](\Lambda) = \left(\zeta \mapsto \frac{d}{dt} G(\Lambda + t\zeta)|_{t=0}\right) \in B(\mathcal{S}).$$

For $\zeta \in \mathcal{S}$ we write $\mathbf{D}[G](\Lambda; \zeta) = \mathbf{D}[G](\Lambda)(\zeta)$ to compress notation.

Proposition 6.1.3. *Let $\mathcal{S}$ be a block algebra and let $\mathcal{D} \subset \mathcal{S}$ be an open set. Let $G : \mathcal{D} \to \mathcal{S}$ be a solution of the SD equation with covariance map $\Phi \in B(\mathcal{S})$. Then for every $\Lambda \in \mathcal{D}$ and $\zeta \in \mathcal{S}$ we have*

$$\begin{aligned} \zeta &= G(\Lambda)^{-1}\mathbf{D}[G](\Lambda; \zeta)G(\Lambda)^{-1} - \Phi(\mathbf{D}[G](\Lambda; \zeta)) \qquad (76) \\ &= \mathbf{D}[G](\Lambda; G(\Lambda)^{-1}\zeta G(\Lambda)^{-1} - \Phi(\zeta)), \\ 0 &= G(\Lambda) + \mathbf{D}[G](\Lambda; \Lambda) + 2\mathbf{D}[G](\Lambda; \Phi(G(\Lambda))). \qquad (77) \end{aligned}$$

Relation (77) plays a key role in proving the bound (23) asserted in Theorem 2.4.6.

Proof. To compress notation further we write $G = G(\Lambda)$ and $G' = \mathbf{D}[G](\Lambda)$. By differentiation of the SD equation we obtain $(\zeta + \Phi(G'(\zeta)))G + (\Lambda + \Phi(G))G'(\zeta) = 0$ and hence $\zeta = G^{-1}G'(\zeta)G^{-1} - \Phi(G'(\zeta))$. Thus the first equality in (76) holds. Now for any linear operators A and B on a finite-dimensional vector space we have $AB = 1 \Rightarrow BA = 1$. Thus $\zeta = G'(G^{-1}\zeta G^{-1} - \Phi(\zeta))$, and hence the second equality in (76) holds. Finally, (77) follows by taking $\zeta = G(\Lambda)$ in (76). $\square$

6.2. **The solution of the SD equation attached to an $\mathcal{S}$-linear form.**

6.2.1. *Definition of $G_L(\Lambda)$.* Fix a block algebra $\mathcal{S}$ and an $\mathcal{S}$-linear form L with Hamel expansion $L = \sum \mathbf{X}_\ell \otimes a_\ell$. We define the set

$$\mathcal{D}_L = \{\Lambda \in \mathcal{S} \mid L(\Xi) - 1_{B(\mathcal{H})} \otimes \Lambda \in (B(\mathcal{H}) \otimes \mathcal{S})^\times\} \subset \mathcal{S}.$$

It is clear that $\mathcal{D}_L$ is nonempty and Lemma 3.1.1 implies that $\mathcal{D}_L$ is open. For $\Lambda \in \mathcal{D}_L$ we put

$$G_L(\Lambda) = \phi^{\mathrm{BF}}_{\mathcal{S}}((L(\Xi) - 1_{B(\mathcal{H})} \otimes \Lambda)^{-1}) \in \mathcal{S},$$

where $\phi^{\mathrm{BF}}_{\mathcal{S}}$ is the $\mathcal{S}$-linear extension of $\phi^{\mathrm{BF}} \in B(\mathcal{H})^\star$. By direct manipulation of series expansions one can verify that $G_L : \mathcal{D}_L \to \mathcal{S}$ is an analytic function. Recall that by Definition 5.2.5 we have $\Phi_L = (\zeta \mapsto \sum a_\ell \zeta a_\ell) \in B(\mathcal{S})$.

Proposition 6.2.2. *The function $G_L : \mathcal{D}_L \to \mathcal{S}$ is a solution of the SD equation with covariance map Φ_L.*

Proof. We specialize Proposition 3.6.4 by taking

$$\mathcal{A} = B(\mathcal{H}) \otimes \mathcal{S}, \ \ \pi = p_{\mathcal{H}} \otimes 1_{\mathcal{S}}, \ \ \rho_\ell = \hat{\Sigma}_\ell \otimes 1_{\mathcal{S}} \quad \text{and}$$
$$A = L(\Xi) - 1_{B(\mathcal{H})} \otimes \Lambda = -1_{B(\mathcal{H})} \otimes \Lambda + \textstyle\sum_\ell \left(\mathrm{i}^\ell \Sigma_\ell \otimes a_\ell + \mathrm{i}^{-\ell} \Sigma_\ell^* \otimes a_\ell\right).$$

To verify that the family $\{\pi\} \cup \{\rho_i\}_{\ell=1}^\infty$ is a Cuntz frame in $\mathcal{A}$ we use (40). To verify that A is quasi-circular we use Lemma 3.4.6. Now in view of (58), the left side of (49) specializes to $p_{\mathcal{H}} \otimes G_L(\Lambda)$ and moreover necessarily $G_L(\Lambda) \in \mathcal{S}^\times$. But we also have

$$\pi A \pi = -p_{\mathcal{H}} \otimes \Lambda, \ \ \pi A \rho_\ell \pi = \mathrm{i}^\ell p_{\mathcal{H}} \otimes a_\ell, \ \ \pi \rho_\ell^* A \pi = \mathrm{i}^{-\ell} p_{\mathcal{H}} \otimes a_\ell,$$

as one verifies by using (38) and (41). Thus the inverse in the algebra $\pi\mathcal{A}\pi$ of the right side of (49) specializes to $-p_{\mathcal{H}} \otimes (\Lambda + \Phi_L(G_L(\Lambda)))$. □

Remark 6.2.3. Proposition 6.2.2 is essentially well-known apart from one small detail. For comparison with a typical proof, see [1, Chap. 5, Secs. 4,5] (main text, not the exercises), and in particular [1, Chap. 5, Lemma 5.5.10]. That proof falls a bit short of proving Proposition 6.2.2 as stated because it relies on an analytic continuation argument to extend a generating function identity proved by combinatorics throughout a connected open set. But we do not know *a priori* that $\mathcal{D}_L$ is connected. (It would be a surprise if it were not but we leave the question aside.) Thus we have presented the operator-theoretic proof of Proposition 6.2.2 suggested by the last exercise in [1] (which does not otherwise seem to be present in the literature in detail) because it makes connectedness of $\mathcal{D}_L$ a non-issue.

Remark 6.2.4. (This is a continuation of the thread of remarks including Remarks 2.6.4 and 5.4.4.) If $(\mathcal{S}, L, \Theta, e)$ is a self-adjoint linearization of $f \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X} \rangle)_{\mathrm{sa}}$, then we have the simple formula

$$S_{\mu_f}(z) = \tau_{\mathcal{S},e}(G_L(\Theta + ze)) \tag{78}$$

for $z \in \mathbb{C} \setminus \operatorname{supp} \mu_f$, which is just a rewrite of (66). This is one way—but not the only way—in which solutions of the SD equation enter the proof of Theorem 2.4.6.

6.3. **Derivatives, symmetries and estimates.** We record some immediate consequences of the construction of $G_L(\Lambda)$ for later use.

6.3.1. *Operator-theoretic representation of the derivative.* By means of the resolvent identity (35) in infinitesimal form, one verifies that

$$\mathbf{D}[G_L](\Lambda; \zeta) = \phi^{\mathrm{BF}}_{\mathcal{S}}((L(\Xi) - 1_{B(\mathcal{H})} \otimes \Lambda)^{-1}(1 \otimes \zeta)(L(\Xi) - 1_{B(\mathcal{H})} \otimes \Lambda)^{-1}), \tag{79}$$

for all $\zeta \in \mathcal{S}$.

6.3.2. *Symmetries.* Note that

$$\Lambda \in \mathcal{D}_L \Leftrightarrow \Lambda^* \in \mathcal{D}_{L^*} \Rightarrow G_L(\Lambda)^* = G_{L^*}(\Lambda^*), \tag{80}$$

$$\Lambda \in \mathcal{D}_L \Leftrightarrow \Lambda^{\mathrm{T}} \in \mathcal{D}_{L^{\mathrm{T}}} \Rightarrow G_L(\Lambda)^{\mathrm{T}} = G_{L^{\mathrm{T}}}(\Lambda^{\mathrm{T}}). \tag{81}$$

Relation (80) holds by the symmetry (56) along with the observation that $*$ commutes with inversion. Relation (81) can be verified by a straightforward calculation exploiting Lemma 5.1.6 and relation (57).

6.3.3. *Estimates.* For any $\mathcal{S}$-linear form L and points $\Lambda, \Lambda_1, \Lambda_2 \in \mathcal{D}_L$, we have estimates

$$[\![G_L(\Lambda)]\!] \leq \left[\!\left[(L(\Xi) - 1_{B(\mathcal{H})} \otimes \Lambda)^{-1}\right]\!\right], \tag{82}$$

$$\begin{aligned}&[\![G_L(\Lambda_1) - G_L(\Lambda_2)]\!]\\ \leq\ &[\![\Lambda_1 - \Lambda_2]\!] \left[\!\left[(L(\Xi) - 1_{B(\mathcal{H})} \otimes \Lambda_1)^{-1}\right]\!\right] \left[\!\left[(L(\Xi) - 1_{B(\mathcal{H})} \otimes \Lambda_2)^{-1}\right]\!\right],\end{aligned} \tag{83}$$

$$[\![\mathbf{D}[G_L](\Lambda)]\!] \leq \left[\!\left[(L(\Xi) - 1_{B(\mathcal{H})} \otimes \Lambda)^{-1}\right]\!\right]^2, \tag{84}$$

$$\begin{aligned}&[\![G_L(\Lambda_1) - G_L(\Lambda_2) - \mathbf{D}[G_L](\Lambda_2; \Lambda_1 - \Lambda_2)]\!]\\ \leq\ &[\![\Lambda_1 - \Lambda_2]\!]^2 \left[\!\left[(L(\Xi) - 1_{B(\mathcal{H})} \otimes \Lambda_1)^{-1}\right]\!\right] \left[\!\left[(L(\Xi) - 1_{B(\mathcal{H})} \otimes \Lambda_2)^{-1}\right]\!\right]^2\end{aligned} \tag{85}$$

which follow directly from the resolvent identity (35), the iterated resolvent identity (36), the estimate (58) and the definitions.

6.4. **The secondary SD equation.** We construct solutions of a secondary form of the Schwinger-Dyson equation by using a variant of the secondary trick. See Proposition 6.4.2 below and its proof.

6.4.1. *The special function $G_{L_1,L_2}(\Lambda_1, \Lambda_2)$.* Let $\mathcal{S}$ be a block algebra. For $j = 1, 2$, let L_j be an $\mathcal{S}$-linear form and let $\Lambda_j \in \mathcal{D}_{L_j}$ be a point. We define

$$G_{L_1,L_2}(\Lambda_1, \Lambda_2) = \phi^{\mathrm{BF}}_{\mathcal{S},\mathcal{S}}((L_1(\Xi) - 1_{B(\mathcal{H})} \otimes \Lambda_1)^{-1}, (L_2(\Xi) - 1_{B(\mathcal{H})} \otimes \Lambda_2)^{-1})$$

where $\phi^{\mathrm{BF}}_{\mathcal{S},\mathcal{S}}$ is the $\mathcal{S}$-bilinear extension of ϕ^{BF}. It is easy to see that $G_{L_1,L_2}(\Lambda_1, \Lambda_2)$ depends analytically on (Λ_1, Λ_2). By Remark 5.3.9 we have

$$[\![G_{L_1,L_2}(\Lambda_1, \Lambda_2)]\!] \leq \left[\!\left[(L_1(\Xi) - 1_{B(\mathcal{H})} \otimes \Lambda_1)^{-1}\right]\!\right] \left[\!\left[(L_2(\Xi) - 1_{B(\mathcal{H})} \otimes \Lambda_2)^{-1}\right]\!\right], \tag{86}$$

which is an estimate analogous to (82).

Proposition 6.4.2. *Let $\mathcal{S}$ be a block algebra. For $j = 1, 2$, let $L_j = \sum \mathbf{X}_\ell \otimes a_{\ell j}$ be an $\mathcal{S}$-linear form and let $\Lambda_j \in \mathcal{D}_{L_j}$ be a point. Then the* secondary SD equation

$$\begin{aligned}&G_{L_1,L_2}(\Lambda_1, \Lambda_2)\\ =\ &\left(\left(\left(G_{L_1}(\Lambda_1)^{-1} \otimes G_{L_2}(\Lambda_2)^{-1} - \sum a_{\ell 1} \otimes a_{\ell 2}\right)^{1 \otimes \mathrm{T}}\right)^{-1}\right)^{1 \otimes \mathrm{T}}\end{aligned} \tag{87}$$

holds. In particular, the expression on the right side is well-defined.

It is worth noting as a consistency check that the expression on the right side remains invariant if we replace the transposition T by any other transposition of $\mathcal{S}$.

Proof. By Remark 5.2.10 it suffices to prove that

$$\zeta = G_{L_1}(\Lambda_1)^{-1} G_{L_1,L_2}(\Lambda_1, \Lambda_2)^\bullet(\zeta) G_{L_2}(\Lambda_2)^{-1} - \sum a_{\ell 1} G_{L_1,L_2}(\Lambda_1, \Lambda_2)^\bullet(\zeta) a_{\ell 2} \tag{88}$$

holds for all $\zeta \in \mathcal{S}$. Let $\mathcal{M}_2$ be a block algebra equipped with a standard basis $\{e_{ij}\}_{i,j=1}^2$. Fix $\zeta \in \mathcal{S}$ arbitrarily and put

$$\Lambda = \Lambda_1 \otimes e_{11} + \Lambda_2 \otimes e_{22} + \zeta \otimes e_{12} \in \mathcal{S} \otimes \mathcal{M}_2.$$

Consider also the $\mathcal{S} \otimes \mathcal{M}_2$-linear form

$$L = L_1 \otimes e_{11} + L_2 \otimes e_{22}.$$

To compress notation put

$$A_j = L_j(\Xi) - 1_{B(\mathcal{H})} \otimes \Lambda_j \in (\mathcal{A} \otimes \mathcal{S})^\times$$

for $j = 1, 2$ and put

$$A = L(\Xi) - 1_{B(\mathcal{H})} \otimes \Lambda \in \mathcal{A} \otimes \mathcal{S} \otimes \mathcal{M}_2.$$

In fact $A \in (\mathcal{A} \otimes \mathcal{S} \otimes \mathcal{M}_2)^\times$, and more precisely

$$A^{-1} = A_1^{-1} \otimes e_{11} + A_2^{-1} \otimes e_{22} + (A_1^{-1}(1_{\mathcal{A}} \otimes \zeta)A_2^{-1}) \otimes e_{12},$$

as one immediately verifies. Thus by the trivial identity (60) we have

$$G_L(\Lambda) = G_{L_1}(\Lambda_1) \otimes e_{11} + G_{L_2}(\Lambda_2) \otimes e_{22} + G_{L_1,L_2}(\Lambda_1, \Lambda_2)^\bullet(\zeta) \otimes e_{12}.$$

By Proposition 6.2.2, the SD equation

$$0 = 1_{\mathcal{S}} \otimes 1_{\mathcal{M}_2} + (\Lambda + \Phi_L(G_L(\Lambda)))G_L(\Lambda)$$

is satisfied. By expanding the right side in the form $\cdots + b \otimes e_{12} + \dots$ we find that

$$\begin{array}{rcl} 0 & = & (\Lambda_1 + \Phi_{L_1}(G_{L_1}(\Lambda_1)))G_{L_1,L_2}(\Lambda_1, \Lambda_2)^\bullet(\zeta) \\ & & +(\zeta + \sum a_{\ell 1} G_{L_1,L_2}(\Lambda_1, \Lambda_2)^\bullet(\zeta) a_{\ell 2})G_{L_2}(\Lambda_2), \end{array}$$

which yields (88) after some further manipulation which we omit. $\square$

Remark 6.4.3. Fix an $\mathcal{S}$-linear form L and a point $\Lambda \in \mathcal{D}_L$. Then we have

$$\mathbf{D}[G_L](\Lambda) = G_{L,L}(\Lambda, \Lambda)^\bullet \tag{89}$$

as one verifies by exploiting the infinitesimal form of the resolvent identity (35). Note that the equation (88) in the case $(L_1, L_2, \Lambda_1, \Lambda_2) = (L, L, \Lambda, \Lambda)$ specializes to the equation (76) obtained through differentiation.

Remark 6.4.4. Fix an $\mathcal{S}$-linear form L and a point $\Lambda \in \mathcal{D}_L$. Let Ψ_L be as in Definition 5.2.5. Recall that if $L = \sum \mathbf{X}_\ell \otimes a_\ell$ is the Hamel expansion of L then $\Psi_L = \sum (-1)^\ell a_\ell^{\otimes 2}$. Then we have

$$\left((G_L(\Lambda)^{-1})^{\otimes 2} - \Psi_L\right)^{-1} = G_{L,L^{\mathrm{T}}}(\Lambda, \Lambda^{\mathrm{T}})^{1 \otimes \mathrm{T}} \tag{90}$$

by the secondary SD equation (87) in the case $(L_1, L_2, \Lambda_1, \Lambda_2) = (L, L^{\mathrm{T}}, \Lambda, \Lambda^{\mathrm{T}})$ along with the symmetry (81). In turn, we have

$$\left[\!\left[((G_L(\Lambda)^{-1})^{\otimes 2} - \Psi_L)^{-1}\right]\!\right] \quad \leq \quad [\![1 \otimes \mathrm{T}]\!] \left[\!\left[(L(\Xi) - 1_{B(\mathcal{H})} \otimes \Lambda)^{-1}\right]\!\right]^2 \tag{91}$$

by Remark 5.3.9, (86), and Lemma 5.1.6.

Remark 6.4.5. Fix a SALT block design $(\mathcal{S}, L, \Theta, e)$ and a point $\Lambda \in \mathcal{D}_L$. We automatically have $\underline{\Lambda} + \Diamond_{\mathcal{S}} \in \mathcal{D}_{\underline{L}}$ and

$$\partial_1 G_{\underline{L}}(\underline{\Lambda} + \Diamond_{\mathcal{S}}) = \mathbf{D}[G_L](\Lambda) \ \text{ and } \ \partial_2 G_{\underline{L}}(\underline{\Lambda} + \Diamond_{\mathcal{S}}) = ((G_L(\Lambda)^{-1})^{\otimes 2} - \Psi_L)^{-1} \tag{92}$$

by (74) and (75) along with (89) and (90).

6.5. **The universal correction** Bias_L^N**.** We first present a general construction needed for the proof of Theorem 2.4.6 which involves solutions of both the "primary" and secondary SD equations. Then in Remark 6.5.5 we specialize the construction to produce our candidate for the correction $\mathrm{bias}_f^N(z)$ figuring in Theorem 2.4.6. Throughout, $\mathcal{S}$ denotes a fixed block algebra.

6.5.1. *A tensor generalization of fourth cumulants.* Let Y be any $\mathcal{S}$-valued random variable such that $\|[\![Y]\!]\|_4 < \infty$ and $\mathbb{E}Y = 0$. Let Z be an independent copy of Y. We define

$$\begin{aligned}\mathbf{C}^{(4)}(Y) \;&=\; \mathbb{E}(Y^* \otimes Y \otimes Y^* \otimes Y) - \mathbb{E}(Y^* \otimes Y \otimes Z^* \otimes Z)\\ &\quad -\mathbb{E}(Y^* \otimes Z \otimes Z^* \otimes Y) - \mathbb{E}(Y^* \otimes Z \otimes Y^* \otimes Z) \in \mathcal{S}^{\otimes 4}.\end{aligned}$$

6.5.2. *Shuffle notation.* For positive integers k we define bilinear maps

$$[\cdot,\cdot]_k : \mathcal{S}^{\otimes k} \times \mathcal{S}^{\otimes k} \to \mathcal{S}^{\otimes 2k},\;\; [x_1 \otimes \cdots \otimes x_k, y_1 \otimes \cdots \otimes y_k]_k = x_1 \otimes y_1 \otimes \cdots \otimes x_k \otimes y_k,$$

$$\langle\cdot,\cdot\rangle_k : \mathcal{S}^{\otimes k} \times \mathcal{S}^{\otimes k} \to \mathcal{S},\;\; \langle x_1 \otimes \cdots \otimes x_k, y_1 \otimes \cdots \otimes y_k\rangle_k = x_1 y_1 \cdots x_k y_k.$$

6.5.3. *Definition of* Bias_L^N. Let $L = \sum \mathbf{X}_\ell \otimes a_\ell$ be any self-adjoint $\mathcal{S}$-linear form. Let $\Lambda \in \mathcal{D}_L$ be a point. To abbreviate notation we write

$$\Phi = \Phi_L \in B(\mathcal{S}),\;\; \Psi = \Psi_L \in \mathcal{S}^{\otimes 2},\;\; X^N = L(\Xi^N) = \sum \Xi_\ell^N \otimes a_\ell \in \mathrm{Mat}_N(\mathcal{S})_{\mathrm{sa}},$$

$$G \;=\; G_L(\Lambda) \in \mathcal{S}^\times,\;\; G' = \mathbf{D}[G_L](\Lambda) \in B(\mathcal{S}),\;\; \check{G} \;=\; \left((G^{-1})^{\otimes 2} - \Psi_L\right)^{-1} \in (\mathcal{S}^{\otimes 2})^\times.$$

By Remark 6.4.4, the object $\check{G}$ above is well-defined. We now define

$$\begin{aligned}(93)\qquad & \widehat{\mathrm{Bias}}_L^N(\Lambda)\\ =\;& \langle[\Psi,\Psi]_2, [\check{G}, G^{\otimes 2}]_2\rangle_4 - \Phi(G)G\\ &+\frac{1}{N}\sum_{i=1}^N \langle \mathbb{E}X^N(i,i)^{\otimes 2}, G^{\otimes 2}\rangle_2 - \frac{1}{N^{3/2}}\sum_{i=1}^N \langle \mathbb{E}X^N(i,i)^{\otimes 3}, G^{\otimes 3}\rangle_3\\ &+\frac{1}{N^2}\sum_{\substack{i,j=1\\ i\neq j}}^N \langle \mathbf{C}^{(4)}(X^N(i,j)), G^{\otimes 4}\rangle_4,\end{aligned}$$

$$(94)\qquad \mathrm{Bias}_L^N(\Lambda) \;=\; G'(\widehat{\mathrm{Bias}}_L^N G^{-1}).$$

The analytic functions

$$\widehat{\mathrm{Bias}}_L^N, \mathrm{Bias}_L^N : \mathcal{D}_L \to \mathcal{S}$$

thus defined we call the *unwrapped universal correction* and *universal correction* indexed by L and N, respectively. We only define the former function to expedite certain calculations—the latter function is the theoretically important one with good symmetry properties. It is a straightforward if tedious matter to verify that Bias_L^N commutes with the C^*-algebra involution just as G_L does. For a constant c independent of N, L and Λ we have

$$(95)\qquad \sup_N [\![\mathrm{Bias}_L^N(\Lambda)]\!] \le c\, [\![(L(\Xi) - 1_{B(\mathcal{H})} \otimes \Lambda)^{-1}]\!]^5$$

by estimates (82), (84) and (91) along with assumption (14).

Remark 6.5.4. Since we are long done with the discussion of Theorems 1.2.4 and 1.2.5 and are focused now on proving Theorem 2.4.6, we feel free to repurpose the letter X in various ways as, for example, in the construction above of Bias_N^L, and later in our discussion of the block Wigner model. This should not cause confusion.

Remark 6.5.5. (This is a continuation of the thread of remarks including Remarks 2.6.4 and 6.2.4.) If $(\mathcal{S}, L, \Theta, e)$ is a self-adjoint linearization of $f \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X} \rangle)_{\mathrm{sa}}$, then our candidate for the correction figuring in Theorem 2.4.6 is defined by the formula

$$\mathrm{bias}_f^N(z) = \tau_{\mathcal{S},e}(\mathrm{Bias}_L^N(\Theta + ze)) \tag{96}$$

for $z \in \mathbb{C} \setminus \operatorname{supp} \mu_f$. This is another distinct way that solutions of the SD equation enter the proof of Theorem 2.4.6.

7. Approximation of solutions of the Schwinger-Dyson equation

We refine a powerful idea from [9] concerning the approximation of solutions of the Schwinger-Dyson equation. See Lemma 7.2.2 below for a short paraphrase of that idea in a simplified geometry. See Proposition 7.3.2 below for the main result of this section, which is an estimate tailored to the proof of Theorem 2.4.6.

7.1. **SD tunnels.**

Definition 7.1.1. Suppose we are given

- a solution $G : \mathcal{D} \to \mathcal{S}$ of the SD equation with covariance map $\Phi \in B(\mathcal{S})$,
- a point $\Lambda_0 \in \mathcal{D}$ and
- (finite) constants $\mathfrak{T} > 0$ and $\mathfrak{G} \geq 1$.

Put

$$\mathcal{T} = \{\Lambda_0 + \mathrm{i}t1_{\mathcal{S}} + \zeta \mid t \in [0,\infty) \text{ and } \zeta \in \mathcal{S} \text{ s.t. } [\![\zeta]\!] \leq 1/\mathfrak{G}\}.$$

Suppose that the following conditions hold:

$$\mathcal{T} \subset \mathcal{D}, \tag{97}$$

$$\sup_{\Lambda \in \mathcal{T}} [\![G(\Lambda)]\!] \leq \mathfrak{G}, \tag{98}$$

$$\sup_{\Lambda \in \mathcal{T}} [\![\mathbf{D}[G](\Lambda)]\!] \leq \mathfrak{G}^2, \tag{99}$$

$$\sup_{\substack{\Lambda,\Lambda' \in \mathcal{T} \\ \text{s.t. } \Lambda \neq \Lambda'}} \frac{[\![G(\Lambda) - G(\Lambda')]\!]}{[\![\Lambda - \Lambda']\!]} \leq \mathfrak{G}^2, \tag{100}$$

$$\sup_{\substack{\Lambda,\Lambda' \in \mathcal{T} \\ \text{s.t. } \Lambda \neq \Lambda'}} \frac{[\![G(\Lambda) - G(\Lambda') - \mathbf{D}[G](\Lambda'; \Lambda - \Lambda')]\!]}{[\![\Lambda - \Lambda']\!]^2} \leq \mathfrak{G}^3, \tag{101}$$

$$\sup_{\Lambda \in \mathcal{T}} [\![G(\Lambda + \mathrm{i}\mathfrak{T}1_{\mathcal{S}})]\!] \leq \frac{1}{2(1 + [\![\Phi]\!])}. \tag{102}$$

In this situation we say that the collection $(G : \mathcal{D} \to \mathcal{S}, \Phi, \Lambda_0, \mathfrak{T}, \mathfrak{G})$ is a *Schwinger-Dyson (SD) tunnel*.

Remark 7.1.2. If $(G : \mathcal{D} \to \mathcal{S}, \Phi, \Lambda_0, \mathfrak{T}, \mathfrak{G})$ is an SD tunnel, then for every $t \in [0, \infty)$, so is $(G : \mathcal{D} \to \mathcal{S}, \Phi, \Lambda_0 + \mathrm{i}t1_{\mathcal{S}}, \mathfrak{T}, \mathfrak{G})$.

Remark 7.1.3. All examples of SD tunnels needed for the proof of Theorem 2.4.6 arise as follows. Let $(\mathcal{S}, L, \Theta, e)$ be a SALT block design. Let c_0, c_1, c_2 and $\mathfrak{T}$ be be the constants from Definition 5.4.1. Put

$$\mathfrak{G}(z) = c_0(1 + [\![L(\Xi)]\!])^{c_1}(1 + 1/\Im z)^{c_2} \tag{103}$$

for $z \in \mathfrak{h}$. We claim that the collection

$$(G_L : \mathcal{D}_L \to \mathcal{S}, \Phi_L, \Theta + ze, \mathfrak{T}, \mathfrak{G}(z)) \tag{104}$$

is an SD tunnel for each fixed $z \in \mathfrak{h}$. To prove the claim, arbitrarily fix $z \in \mathfrak{h}$, $t \in [0, \infty)$ and $\zeta \in \mathcal{S}$ such that $[\![\zeta]\!] \leq 1/\mathfrak{G}(z)$, and put

$$\Lambda = \Theta + ze + \mathrm{i}t1_{\mathcal{S}} + \zeta.$$

We then have

$$\Lambda \in \mathcal{D}_L \ \text{ and } \ \left[\!\left[(L(\Xi) - 1_{B(\mathcal{H})} \otimes \Lambda)^{-1}\right]\!\right] \leq \begin{cases} \mathfrak{G}(z) & \text{in general,} \\ \frac{1}{2(1+[\![\Phi]\!])} & \text{for } t \geq \mathfrak{T}, \end{cases}$$

by Definition 5.4.1 and Remark 5.4.5. In particular, (104) satisfies (97) for each fixed $z \in \mathfrak{h}$. In turn, it follows by (82) that (104) satisfies (98) and (102) for each fixed $z \in \mathfrak{h}$. Given also $\Lambda' \in \mathcal{D}_L$ with "primed" variables, we have

$$[\![G_L(\Lambda) - G_L(\Lambda')]\!] \leq [\![\Lambda - \Lambda']\!]\,\mathfrak{G}(z)\mathfrak{G}(z'), \quad [\![\mathbf{D}[G_L](\Lambda)]\!] \leq \mathfrak{G}(z)^2,$$
$$[\![G_L(\Lambda) - G_L(\Lambda') - \mathbf{D}[G_L](\Lambda'; \Lambda - \Lambda')]\!] \leq [\![\Lambda - \Lambda']\!]^2\,\mathfrak{G}(z)\mathfrak{G}(z')^2$$

by (83), (84) and (85), respectively. It follows that (104) also satisfies (98), (99) and (100) for each fixed $z \in \mathfrak{h}$. The claim is proved. We note also that we have a bound

$$\left[\!\left[\mathrm{Bias}_L^N(\Lambda)\right]\!\right] \leq c\mathfrak{G}(z)^5 \tag{105}$$

for a constant c independent of L, N and z by (95). This last estimate turns out to be the crucial point for proving the bound (24) asserted in Theorem 2.4.6.

Remark 7.1.4. Definition 7.1.1 is not particularly delicate or economical. Indeed, conditions (97), (98) and (102) alone imply that $(G : \mathcal{D} \to \mathcal{S}, \Phi, \Lambda_0, \mathfrak{T}, c\mathfrak{G})$ is an SD tunnel, where $c > 1$ is an absolute constant. However, for the present purpose, no advantage is gained by reformulating Definition 7.1.1 in more economical fashion since all the properties (97)—(102) are needed to prove Proposition 7.3.2 below.

7.2. **The key lemma.** Before working out our main estimate, we first prove a simple lemma to explain the mechanism by which SD tunnels control errors. The lemma captures a key idea of [9] but works with a simpler geometry. The lemma uses only the first and last of the defining conditions of an SD tunnel.

7.2.1. *Setup for the key lemma.*

- Let $(G : \mathcal{D} \to \mathcal{S}, \Phi, \Lambda_0, \mathfrak{T}, \mathfrak{G})$ satisfy conditions (97) and (102) of the definition of an SD tunnel.
- Let $F = (t \mapsto F_t) : [0, \mathfrak{T}] \to \mathcal{S}$ be a continuous function.

For $t \in [0, \mathfrak{T}]$ we put

$$\Lambda_t = \Lambda_0 + \mathrm{i}t1_{\mathcal{S}}, \ \ G_t \ = \ G(\Lambda_t), \ \ E_t = 1_{\mathcal{S}} + (\Lambda_t + \Phi(F_t))F_t.$$

Note that we have $\Lambda_t \in \mathcal{D}$ by definition of an SD tunnel and hence G_t is well-defined. In turn we define constants

$$\mathfrak{C}_0 = 2(1 + [\![\Phi]\!]), \ \ \mathfrak{F} \ = \ 1 \vee \sup_{t\in[0,\mathfrak{T}]} [\![F_t]\!]\,, \ \ \mathfrak{A} = \sup_{t\in[0,\mathfrak{T}]} [\![E_t]\!]\,.$$

The quantity $\mathfrak{A}$ is a natural measure of the failure of F to satisfy the SD equation. We emphasize that we assume nothing of the function F beyond continuity.

Lemma 7.2.2. *If*

$$\mathfrak{C}_0\mathfrak{G}\mathfrak{F}\mathfrak{A} < 1 \quad \textit{and} \tag{106}$$

$$[\![F_{\mathfrak{T}}]\!] < 1, \tag{107}$$

then for every $t \in [0,\mathfrak{T}]$*, the inverse* $H_t = -(\Lambda_t + \Phi(F_t))^{-1}$ *exists,*

$$[\![H_t]\!] \;\leq\; 2\,[\![F_t]\!]\,, \tag{108}$$

$$[\![\Phi(H_tE_t)]\!] \leq 1/\mathfrak{G}, \quad \textit{(hence)}\ \Lambda_t - \Phi(H_tE_t) \in \mathcal{D} \ \textit{and} \tag{109}$$

$$H_t - H_tE_t = F_t = G(\Lambda_t - \Phi(H_tE_t)) - H_tE_t. \tag{110}$$

Proof. Fix $t \in [0,\mathfrak{T}]$ arbitrarily. Hypothesis (106) implies that $[\![E_t]\!] \leq 1/2$. By Lemma 3.1.1 it follows that H_t is well-defined and satisfies (108). Then claim (109) holds by (97), (106) and (108). It remains only to prove claim (110), and since the first equality in (110) holds by definition of H_t, we have only to prove the second equality. By the Weierstrass Approximation Theorem, we may assume that F depends polynomially and *a fortiori* analytically on t. Put

$$\widehat{H}_t = G(\Lambda_t - \Phi(H_tE_t)) \ \text{ and } \ \widehat{F}_t = \widehat{H}_t - H_tE_t.$$

Note that $\widehat{F}_t$ depends analytically on t. It is enough to prove $F_t \equiv \widehat{F}_t$. In any case, since G satisfies the SD equation with covariance map Φ, we have

$$1_{\mathcal{S}} + (\Lambda_t - \Phi(H_tE_t) + \Phi(\widehat{H}_t))\widehat{H}_t = 1_{\mathcal{S}} + (\Lambda_t + \Phi(\widehat{F}_t))\widehat{H}_t = 0$$

and hence $\widehat{H}_t = -(\Lambda_t + \Phi(\widehat{F}_t))^{-1}$. We thus have

$$F_t - \widehat{F}_t = H_t - \widehat{H}_t = H_t\Phi(F_t - \widehat{F}_t)\widehat{H}_t = H_t\Phi(F_t - \widehat{F}_t)G(\Lambda_t - \Phi(H_tE_t)),$$

where at the second step we use the resolvent identity (35). Finally, by (102), (107), (108) and (109) we have

$$[\![H_{\mathfrak{T}}]\!]\,[\![\Phi]\!]\,[\![G(\Lambda_{\mathfrak{T}} - \Phi(H_{\mathfrak{T}}E_{\mathfrak{T}}))]\!] < 1,$$

hence the difference $F_t - \widehat{F}_t$ vanishes identically for t near $\mathfrak{T}$ and hence $F_t \equiv \widehat{F}_t$ by analytic continuation. $\square$

Remark 7.2.3. We work out the simplest concrete example of the phenomenon described by the lemma. Let $\sigma : \mathbb{C} \setminus [-2,2] \to \mathbb{C}$ be the Stieltjes transform of the semicircle law. As is well-known, $\sigma(z)$ is the unique bounded analytic solution of the equation $1 + (z + \sigma(z))\sigma(z) = 0$ in the domain $\mathbb{C} \setminus [-2,2]$. Now fix $z_0 \in \mathfrak{h}$ arbitrarily. It is easy to see that

$$(G : \mathcal{D} \to \mathcal{S}, \Phi, \Lambda_0, \mathfrak{T}, \mathfrak{G}) = \left(\sigma : \mathbb{C} \setminus [-2,2] \to \mathbb{C}, 1, z_0, 4, 1 \vee \frac{2}{\Im z_0}\right)$$

satisfies conditions (97) and (102) of the definition of an SD tunnel. Now fix a continuous function $(t \mapsto F_t) : [0,4] \to \mathbb{C}$ and put $\mathfrak{F} = 1 \vee \sup_{t\in[0,4]} |F_t|$. In turn put $E_t = 1 + (z_0 + \mathrm{i}t + F_t)F_t$ for $t \in [0,4]$ and $\mathfrak{A} = \sup_{t\in[0,4]} |E_t|$. Assume that $|F_4| < 1$ and $\mathfrak{F}\mathfrak{A} < \frac{1}{4}\left(1 \wedge \frac{\Im z_0}{2}\right)$. Then by Lemma 7.2.2 we have $H_0 = -(z_0 + F_0)^{-1} \neq \infty$, $|H_0| \leq 2|F_0|$, $|H_0E_0| \leq 1 \wedge \frac{\Im z_0}{2}$ and finally $F_0 = \sigma(z_0 - H_0E_0) - H_0E_0$. This last equation is at first glance a bit strange but in fact the strategy of writing $\rho(z) = \sigma(z - \delta) - \delta$ to estimate $\rho(z) - \sigma(z)$ has long been in use. See for example [2, Eq. 4.11].

Remark 7.2.4. Equation (110) is not an obvious target to shoot for! But once (110) is written down, it is clear that it offers excellent opportunities for systematically estimating the difference $[\![F_0 - G(\Lambda_0)]\!]$. This surprising and powerful idea we learned from [9]. The importance and utility of this idea cannot be overestimated.

7.3. **The tunnel estimates.** We now use Lemma 7.2.2 to obtain an estimate in terms of parameters over which we will be able to gain good control. In particular, the estimate is designed to take advantage of Remark 5.4.6 above.

7.3.1. *Setup for the tunnel estimates.*

- Let $(G : \mathcal{D} \to \mathcal{S}, \Phi, \Lambda_0, \mathfrak{T}, \mathfrak{G})$ be an SD tunnel.
- Let $\mathfrak{L} \in [1, \infty)$ be a constant.
- Let $F = (t \mapsto F_t) : [0, \infty) \to \mathcal{S}$ be a Lipschitz continuous function with Lipschitz constant bounded by $\mathfrak{L}$ and satisfying $\sup_{t\in[\mathfrak{T},\infty)} [\![F_t]\!] \leq 1/2$.

For $t \in [0, \infty)$ we put

$$\Lambda_t = \Lambda_0 + \mathrm{i}t1_{\mathcal{S}},\ \ E_t = 1_{\mathcal{S}} + (\Lambda_t + \Phi(F_t))F_t,$$

and we define constants

$$\mathfrak{C} = 99e^{2\mathfrak{T}}(1 + [\![\Lambda_0]\!] + [\![\Phi]\!]),\ \ \mathfrak{E} = \frac{1}{2}\,[\![E_0]\!] + \frac{1}{2}\int_0^\infty [\![E_t]\!]\, e^{-t}dt.$$

The integral converges since $[\![E_t]\!]$ has at worst linear growth as $t \to \infty$.

Proposition 7.3.2. *Data, notation and assumptions are as above. We have*

$$[\![F_0 - G(\Lambda_0)]\!] \leq (\mathfrak{C}\mathfrak{G}\mathfrak{L})^6(\mathfrak{E} + \mathfrak{E}^2), \tag{111}$$

$$\left[\!\!\left[F_0 + \mathbf{D}[G](\Lambda_0; E_0G(\Lambda_0)^{-1}) - G(\Lambda_0)\right]\!\!\right] \leq (\mathfrak{C}\mathfrak{G}\mathfrak{L})^{12}(\mathfrak{E}^2 + \mathfrak{E}^4). \tag{112}$$

The exponents of $\mathfrak{C}$, $\mathfrak{G}$ and $\mathfrak{L}$ are of no importance in the sequel. They could be replaced by any larger absolute constants without disturbing later arguments. Only the exponents of $\mathfrak{E}$ will be important.

Proof. In anticipation of applying Lemma 7.2.2, we put

$$\mathfrak{F} \;=\; 1 \vee \sup_{t\in[0,\mathfrak{T}]} [\![F_t]\!] = 1 \vee \sup_{t\in[0,\infty)} [\![F_t]\!] \leq \sqrt{\mathfrak{C}}\mathfrak{L}, \tag{113}$$

$$\mathfrak{A} \;=\; \sup_{t\in[0,\mathfrak{T}]} [\![E_t]\!] \leq \mathfrak{C}\mathfrak{F}^2. \tag{114}$$

We have also noted here some crude bounds needed later. Now write

$$G_0 = G(\Lambda_0),\ \ G_0' = \mathbf{D}[G](\Lambda_0),\ \ V_0 = G_0'(E_0G_0^{-1}) = G_0'(\Phi(G_0E_0)) + G_0E_0$$

in order to abbreviate notation. The last equality above is an instance of (76).

We now claim that

$$[\![F_0 - G_0]\!] \leq \mathfrak{C}\mathfrak{G}^2\mathfrak{F}([\![E_0]\!] + \mathbf{1}_{\mathfrak{C}\mathfrak{G}\mathfrak{F}\mathfrak{A}\geq 1}), \tag{115}$$

$$[\![F_0 + V_0 - G_0]\!] \leq \mathfrak{C}^2\mathfrak{G}^5\mathfrak{F}^2([\![E_0]\!]^2 + \mathbf{1}_{\mathfrak{C}\mathfrak{G}\mathfrak{F}\mathfrak{A}\geq 1}). \tag{116}$$

If $\mathfrak{C}\mathfrak{G}\mathfrak{F}\mathfrak{A} \geq 1$, then crude estimates based on the definition of an SD tunnel along with the bound (114) suffice. We may therefore assume without loss of generality that $\mathfrak{C}\mathfrak{G}\mathfrak{F}\mathfrak{A} < 1$, in which case the hypotheses of Lemma 7.2.2 are fulfilled. Thus it follows via (100), (108), (109) and (110) that

$$[\![F_0 - G_0]\!] \leq [\![H_0 - G_0]\!] + [\![H_0E_0]\!] \leq \mathfrak{G}^2(2\,[\![\Phi]\!]\,\mathfrak{F}\,[\![E_0]\!]) + 2\mathfrak{F}\,[\![E_0]\!]\,.$$

Thus the claim (115) is proved. To prove (116), we begin by noting the identity

$$\begin{aligned} F_0 + V_0 - G_0 &= G(\Lambda_0 - \Phi(H_0 E_0)) - G_0 + G_0'(\Phi(H_0 E_0)) \\ &\quad +(G_0 - H_0)E_0 + G_0'(\Phi((G_0 - H_0)E_0)) \end{aligned}$$

derived from (110). Then, reasoning as in the proof of (115), but now also using (101) and (102), we find that

$$\begin{aligned} &[\![F_0 + V_0 - G_0]\!] \\ \leq\ & \mathfrak{G}^3(2\,[\![\Phi]\!]\,\mathfrak{F}\,[\![E_0]\!])^2 + \mathfrak{G}^2(2\,[\![\Phi]\!]\,\mathfrak{F}\,[\![E_0]\!])\,[\![E_0]\!] + \mathfrak{G}^3\,[\![\Phi]\!]\,(\mathfrak{G}^2(2\,[\![\Phi]\!]\,\mathfrak{F}\,[\![E_0]\!]))\,[\![E_0]\!]\,. \end{aligned}$$

Thus the claim (116) is proved.

We next claim that

$$\mathfrak{A} \leq \sqrt{\mathfrak{C}\mathfrak{F}\mathfrak{L}}(\sqrt{\mathfrak{E}} + \mathfrak{E}). \tag{117}$$

To prove the claim, consider the function $b(t) = e^{-t}\,[\![E_t]\!]$ defined for $t \in [0,\infty)$. Since b is continuous and tends to 0 at infinity, b achieves its maximum at some point $t_0 \in [0,\infty)$. Clearly, we have

$$b(t_0)e^{\mathfrak{T}} \geq \mathfrak{A},\ \ 2\mathfrak{E} \geq \int_0^\infty b(t)dt.$$

Now fix $t > t_0$ arbitrarily. We have

$$\begin{aligned} |b(t_0) - b(t)| &\leq e^{-t_0}\,[\![E_{t_0}]\!]\,(1 - e^{t_0 - t}) + e^{-t}\,[\![E_{t_0} - E_t]\!] \\ &\leq \big(b(t_0) + (1 + [\![\Phi]\!]\,\mathfrak{L})\mathfrak{F} + ([\![\Lambda_0]\!] + e^{-t}t + [\![\Phi]\!]\,\mathfrak{F})\mathfrak{L}\big)\,|t_0 - t| \\ &\leq \left(b(t_0) + \frac{\mathfrak{C}\mathfrak{F}\mathfrak{L}}{8e^{2\mathfrak{T}}}\right)|t_0 - t|. \end{aligned}$$

Thus there exists a right triangle with altitude $b(t_0)$ and base of length $\frac{b(t_0)}{b(t_0) + \frac{\mathfrak{C}\mathfrak{F}\mathfrak{L}}{8e^{2\mathfrak{T}}}}$ under the graph of b. Now in general for $K_1, x \geq 0$ and $K_2 > 0$ we have

$$4K_1 \geq \frac{x^2}{x + K_2} \Rightarrow x \leq \sqrt{8K_1K_2} + 8K_1.$$

The claim (117) now follows after some further trivial manipulations which we omit.

Finally, by combining (113), (115), (116) and (117) we obtain bounds

$$\begin{aligned} [\![F_0 - G_0]\!] &\leq \mathfrak{C}^{3/2}\mathfrak{G}^2\mathfrak{L}(2\mathfrak{E} + \mathbf{1}_{\mathfrak{C}^2\mathfrak{G}^{3/2}\mathfrak{L}^{3/2}(\sqrt{\mathfrak{E}}+\mathfrak{E})\geq 1}), \\ [\![F_0 + V_0 - G_0]\!] &\leq \mathfrak{C}^3\mathfrak{G}^5\mathfrak{L}^2(4\mathfrak{E}^2 + \mathbf{1}_{\mathfrak{C}^2\mathfrak{G}^{3/2}\mathfrak{L}^{3/2}(\sqrt{\mathfrak{E}}+\mathfrak{E})\geq 1}), \end{aligned}$$

whence the result after using Chebychev bounds and simplifying brutally. □

8. Matrix identities

Throughout this section we fix a block algebra $\mathcal{S}$. Working in a purely algebraic setting, we build up a catalog of identities satisfied by finite chunks of an infinite matrix with entries in $\mathcal{S}$. The identities are chosen to illuminate the structure of random matrices of the form (33) and are a further contribution to our stock of tools for concentration. All the identities derived here are meaningful in the case $\mathcal{S} = \mathbb{C}$, in which case many of these identities are familiar from the study of resolvents of Wigner matrices.

8.1. **An *ad hoc* infinite matrix formalism.** When we write $\mathrm{Mat}_{k\times\ell}(\mathcal{S})$, we now allow k or ℓ or both to be infinite, in which case we mean for the corresponding matrix indices to range over all positive integers. Addition, multiplication and adjoints of (possibly) infinite matrices are defined as before, although we never attempt to multiply such matrices unless one of them has only finitely many nonzero entries. For each integer $N>0$, let $\mathcal{I}_N$ denote the family of nonempty subsets of the set $\{1,\dots,N\}$. Given a finite nonempty set $I=\{i_1<\cdots<i_k\}$ of positive integers, let $\mathbf{f}_I\in\mathrm{Mat}_{k\times\infty}(\mathcal{S})$ and $\mathbf{e}_I\in\mathrm{Mat}_\infty(\mathcal{S})$ be defined by

$$\mathbf{f}_I(i,j)=\sum_{\alpha=1}^k \mathbf{1}_{(i,j)=(\alpha,i_\alpha)}1_{\mathcal{S}}\ \text{ and }\ \mathbf{e}_I(i,j)=\sum_{\alpha=1}^k \mathbf{1}_{(i,j)=(i_\alpha,i_\alpha)}1_{\mathcal{S}},$$

respectively. Note that $\mathbf{f}_I\mathbf{f}_I^*=\mathbf{I}_{|I|}\otimes 1_{\mathcal{S}}$ and $\mathbf{f}_I^*\mathbf{f}_I=\mathbf{e}_I$, where $|I|$ denotes the cardinality of I. Note that for all $A\in\mathrm{Mat}_\infty(\mathcal{S})$ and finite sets I and J of positive integers, the finite matrix $\mathbf{f}_IA\mathbf{f}_J^*\in\mathrm{Mat}_{|I|\times|J|}(\mathcal{S})$ is the result of striking all rows of A with indices not in I and all columns of A with indices not in J. Thus the matrices $\mathbf{f}_I$ allow us to pick out finite chunks of a matrix $A\in\mathrm{Mat}_\infty(\mathcal{S})$ and to use the familiar rules of matrix algebra itself to manipulate the chunks. For $A\in\mathrm{Mat}_\infty(\mathcal{S})$ with only finitely many nonzero entries, we define $\mathrm{tr}_{\mathcal{S}}A=\sum_i A(i,i)$. For such A we also define $[\![A]\!]=[\![\mathbf{f}_IA\mathbf{f}_I^*]\!]$ for any finite set I of positive integers such that $\mathbf{e}_IA\mathbf{e}_I=A$, which is independent of I. For each $\zeta\in\mathcal{S}$, let $\mathbf{I}_\infty\otimes\zeta\in\mathrm{Mat}_\infty(\mathcal{S})$ denote the infinite diagonal matrix with diagonal entries ζ.

8.2. **The setup for studying matrix identities.**

8.2.1. *Data and assumption.* We fix a triple (X,Λ,Φ) where

- $X\in\mathrm{Mat}_\infty(\mathcal{S})$,
- $\Lambda\in\mathcal{S}$ and
- $\Phi\in B(\mathcal{S})$,

subject to the condition

$$\mathbf{f}_I\left(\frac{X}{\sqrt{N}}-\mathbf{I}_\infty\otimes\Lambda\right)\mathbf{f}_I^*\in\mathrm{GL}_{|I|}(\mathcal{S})\ \text{ for } N \text{ and } I\in\mathcal{I}_N. \tag{118}$$

Here and below N is understood to range over the positive integers. Below we will define and analyze various functions of the triple (X,Λ,Φ), calling them *recipes*.

Remark 8.2.2. All triples (X,Λ,Φ) needed to prove Theorem 2.4.6 arise as follows. Let $(\mathcal{S},L,\Theta,e)$ be any SALT block design. Let $\bigcup L(\Xi^N)\in\mathrm{Mat}_\infty(\mathcal{S})_{\mathrm{sa}}$ denote the infinite matrix gotten by cobbling together the matrices $L(\Xi^N)\in\mathrm{Mat}_N(\mathcal{S})$ for varying N using assumption (16). Let $\mathbf{z}$ be as in Theorem 2.4.6. Let $\mathbf{t}$ be a real random variable independent of $\sigma(\mathcal{F},\mathbf{z})$ which with probability 1/2 is concentrated at the origin and with probability 1/2 is standard exponential. (The motivation for using the random variable $\mathbf{t}$ comes from Proposition 7.3.2 above.) Let $\Phi_L\in B(\mathcal{S})$ be as in Definition 5.2.5. Then the triple

$$(X,\Lambda,\Phi)=\left(\bigcup L(\Xi^N),\Theta+\mathbf{z}e+\mathrm{i}\mathbf{t}1_{\mathcal{S}},\Phi_L\right) \tag{119}$$

satisfies (118) with probability 1 by Remark 5.4.6.

8.2.3. *The first group of recipes.* For N and $I \in \mathcal{I}_N$ we define

$$\begin{aligned}
R_I^N &= \mathbf{f}_I^* \left(\mathbf{f}_I \left(\frac{X}{\sqrt{N}} - \mathbf{I}_\infty \otimes \Lambda\right) \mathbf{f}_I^*\right)^{-1} \mathbf{f}_I \in \mathrm{Mat}_\infty(\mathcal{S}), \quad F_I^N = \frac{1}{N}\mathrm{tr}_{\mathcal{S}} R_I^N \in \mathcal{S},\\
T_I^N &= \left(\zeta \mapsto \frac{1}{N}\sum_{i,j\in I} R_I^N(i,j)\zeta R_I^N(j,i)\right) \in B(\mathcal{S}),\\
U_I^N &= \frac{1}{N}\sum_{i,j\in I} R_I^N(i,j)^{\otimes 2} \in \mathcal{S}^{\otimes 2}.
\end{aligned}$$

Note that R_I^N is well-defined by assumption (118). For N put

$$\mathcal{I}_N^{(2)} = \{(I,J) \in \mathcal{I}_N \times \mathcal{I}_N \mid J \subset I,\, I \setminus J \in \mathcal{I}_N,\, |J| \leq 2\}.$$

For N and $(I,J) \in \mathcal{I}_N^{(2)}$ put

$$R_{I,J}^N = \mathbf{f}_J R_I^N \mathbf{f}_J^* \in \mathrm{Mat}_{|J|}(\mathcal{S}).$$

The recipes in the first group do not depend on Φ, whereas the remaining recipes we are about to define do depend on Φ.

8.2.4. *Recipes of the second group.* For N and $I \in \mathcal{I}_N$ put

$$\begin{aligned}
E_I^N &= 1_{\mathcal{S}} + (\Lambda + \Phi(F_I^N))F_I^N \in \mathcal{S},\\
H_I^N &= \begin{cases} -(\Lambda + \Phi(F_I^N))^{-1} \in \mathcal{S}^\times & \text{if } \left[\!\left[E_I^N\right]\!\right] < 1/2,\\ 0 \in \mathcal{S} & \text{if } \left[\!\left[E_I^N\right]\!\right] \geq 1/2. \end{cases}
\end{aligned}$$

Note that H_I^N is well-defined by Lemma 3.1.1. For N, $(I,J) \in \mathcal{I}_N^{(2)}$ and $j_1, j_2 \in J$, we define

$$\begin{aligned}
H_{I,J}^N &= \mathbf{I}_{|J|} \otimes H_{I\setminus J}^N \in \mathrm{Mat}_{|J|}(\mathcal{S}),\\
\frac{Q_{I,J}^N}{\sqrt{N}} &= -\frac{\mathbf{f}_J X \mathbf{f}_J^*}{\sqrt{N}} + \frac{\mathbf{f}_J X R_{I\setminus J}^N X \mathbf{f}_J^*}{N} - \mathbf{I}_{|J|} \otimes \Phi(F_{I\setminus J}^N) \in \mathrm{Mat}_{|J|}(\mathcal{S}),\\
Q_{I,J,j_1,j_2}^N &= \mathbf{f}_{j_1}\mathbf{f}_J^* Q_{I,J}^N \mathbf{f}_J \mathbf{f}_{j_2}^* \in \mathcal{S},\\
\frac{P_{I,J}^N}{\sqrt{N}} &= \left(A \mapsto \frac{1}{N}\mathrm{tr}_{\mathcal{S}}(R_{I\setminus J}^N X \mathbf{f}_J^* A \mathbf{f}_J X R_{I\setminus J}^N) - T_{I\setminus J}^N \circ \Phi \circ \mathrm{tr}_{\mathcal{S}}(A)\right)\\
&\qquad \in B(\mathrm{Mat}_{|J|}(\mathcal{S}), \mathcal{S}),\\
P_{I,J,j_1,j_2}^N &= (\zeta \mapsto P_{I,J}^N(\mathbf{f}_J \mathbf{f}_{j_1}^* \zeta \mathbf{f}_{j_2} \mathbf{f}_J^*)) \in B(\mathcal{S}),\\
\Delta_{I,J}^N &= H_{I,J}^N Q_{I,J}^N + \sqrt{N}\mathbf{I}_{|J|}\mathbf{1}_{\left[\!\left[E_{I\setminus J}^N\right]\!\right] \geq 1/2} \in \mathrm{Mat}_{|J|}(\mathcal{S}).
\end{aligned}$$

8.2.5. *Abuses of notation.* We write

$$\Delta^k R_{I,J}^N = (\Delta_{I,J}^N)^k R_{I,J}^N \text{ and } \Delta R_{I,J}^N = \Delta^1 R_{I,J}^N.$$

We often write j where we should more correctly write $\{j\}$, e.g., we write $Q_{I,j}^N$ instead of $Q_{I,\{j\}}^N$. Note that

$$R_{I,j}^N = R_I^N(j,j), \quad H_{I,j}^N = H_{I\setminus j}^N, \quad Q_{I,j,j,j}^N = Q_{I,j}^N, \quad P_{I,j,j,j}^N = P_{I,j}^N.$$

In the same spirit, we occasionally write N in place of $\{1,\ldots,N\}$.

Remark 8.2.6. This is a continuation of Remark 8.2.2 and furthermore a continuation of the thread of remarks including Remarks 2.6.4, 5.4.4 and 6.2.4. Suppose now that $(\mathcal{S}, L, \Theta, e)$ is a self-adjoint linearization of some $f \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X} \rangle)_{\mathrm{sa}}$. Then with $\tau_{\mathcal{S},e}$ as in Definition 5.2.6, we have

$$S_{\mu_f^N}(\mathbf{z}) = \tau_{\mathcal{S},e}(F_N^N) \text{ on the event } \mathbf{t} = 0. \tag{120}$$

This representation of $S_{\mu_f^N}(\mathbf{z})$ is just a rewrite of (65). Next, let $(\underline{\mathcal{S}}, \underline{L}, \underline{\Theta} + \Diamond_{\mathcal{S}}, \underline{e})$ be the "underlined" SALT block design derived from $(\mathcal{S}, L, \Theta, e)$ via Lemma 5.5.4. Consider the triple

$$(\underline{X}, \underline{\Lambda}, \underline{\Phi}) = \left(\bigcup \underline{L}(\Xi^N), \underline{\Theta} + \Diamond_{\mathcal{S}} + \mathbf{z}\underline{e} + \mathrm{i}\mathbf{t}1_{\underline{\mathcal{S}}}, \Phi_{\underline{L}}\right), \tag{121}$$

which again satisfies assumption (118) with probability 1. Let the recipes attached to the underlined triple $(\underline{X}, \underline{\Lambda}, \underline{\Phi})$ be denoted with underlines in order to distinguish them from those attached to the triple (X, Λ, Φ) defined by (119). We then have

$$\partial_1 \underline{F}_I^N = T_I^N \text{ and } \partial_2 \underline{F}_I^N = U_I^N \tag{122}$$

by (61), (62), (74) and (75). The relations (122) joined with the relations (92) will be crucial for the proof of relation (25) of Theorem 2.4.6.

Remark 8.2.7. The recipe U_I^N does not figure in any identities stated in §8 but does become an important random variable later. We therefore include its definition here so that §8.2 can serve as a handy catalog of the basic random variables.

Remark 8.2.8. Note that R_I^N is the inverse of the matrix $\mathbf{e}_I\left(\frac{X}{\sqrt{N}} - \mathbf{I}_\infty \otimes \Lambda\right)\mathbf{e}_I$ as computed in the algebra $\mathbf{e}_I \mathrm{Mat}_\infty(\mathcal{S})\mathbf{e}_I$ the identity element of which is $\mathbf{e}_I$. This observation simplifies calculations below on several occasions.

8.3. **Basic identities.** We obtain block-type generalizations of matrix identities familiar from the study of resolvents of Wigner matrices.

Lemma 8.3.1. *For N and $I \in \mathcal{I}_N$, along with any positive integer k,*

$$R_I^{N+1} = R_I^N + \sum_{\nu=1}^{k-1}\left(\delta_N R_I^N \frac{\mathbf{e}_I X \mathbf{e}_I^*}{\sqrt{N}}\right)^\nu R_I^N + \left(\delta_N R_I^N \frac{\mathbf{e}_I X \mathbf{e}_I^*}{\sqrt{N}}\right)^k R_I^{N+1}, \tag{123}$$

where $\delta_N = \sqrt{N}\left(\frac{1}{\sqrt{N}} - \frac{1}{\sqrt{N+1}}\right)$.

Proof. By induction we may assume $k = 1$. Then, in view of Remark 8.2.8, formula (123) is merely an instance of the resolvent identity (35). ∎

Lemma 8.3.2. *For N and $(I, J) \in \mathcal{I}_N^{(2)}$,*

$$R_{I,J}^N = \left(\frac{\mathbf{f}_J X \mathbf{f}_J^*}{\sqrt{N}} - \mathbf{I}_{|J|} \otimes \Lambda - \frac{\mathbf{f}_J X R_{I\setminus J}^N X \mathbf{f}_J^*}{N}\right)^{-1}, \tag{124}$$

$$R_I^N - R_{I\setminus J}^N = \left(\mathbf{f}_J^* - R_{I\setminus J}^N \frac{X}{\sqrt{N}}\mathbf{f}_J^*\right) R_{I,J}^N \left(\mathbf{f}_J - \mathbf{f}_J \frac{X}{\sqrt{N}} R_{I\setminus J}^N\right). \tag{125}$$

In particular, we automatically have $R_{I,J}^N \in \mathrm{GL}_{|J|}(\mathcal{S})$.

Proof. In Proposition 3.5.2, let us now take

$$\mathcal{A} = \mathbf{e}_I \mathrm{Mat}_\infty(\mathcal{S})\mathbf{e}_I,\ \ x = \mathbf{e}_I\left(\frac{X}{\sqrt{N}} - \mathbf{I}_\infty \otimes \Lambda\right)\mathbf{e}_I,\ \ \pi = \mathbf{e}_J,\ \ \pi^\perp = \mathbf{e}_{I\setminus J},\ \ \sigma = \mathbf{e}_I,$$

and again let us take advantage of Remark 8.2.8. Rewritten in the form

$$\mathbf{e}_J R_I^N \mathbf{e}_J = \mathbf{f}_J^* R_{I,J}^N \mathbf{f}_J = \mathbf{f}_J^*\left(\mathbf{f}_J\left(x - xR_{I\setminus J}^N x\right)\mathbf{f}_J^*\right)^{-1}\mathbf{f}_J,$$

identity (124) becomes a special case of (43). Similarly, rewritten in the form

$$R_I^N - R_{I\setminus J}^N = \left(\mathbf{e}_J - R_{I\setminus J}^N x\mathbf{e}_J\right) R_I^N \left(\mathbf{e}_J - \mathbf{e}_J x R_{I\setminus J}^N\right),$$

identity (125) becomes a specialization of (44). □

Lemma 8.3.3. *For N and $(I,J) \in \mathcal{I}_N^{(2)}$, along with any positive integer k,*

$$R_{I,J}^N = H_{I,J}^N + \sum_{\nu=1}^{k-1}\frac{(H_{I,J}^N Q_{I,J}^N)^\nu H_{I,J}^N}{N^{\nu/2}} + \frac{\Delta^k R_{I,J}^N}{N^{k/2}}. \tag{126}$$

Proof. By induction on k we may assume $k = 1$. Rewrite (124) in the form

$$-\left(\mathbf{I}_{|J|} \otimes (\Lambda + \Phi(F_{I\setminus J}^N))\right)R_{I,J}^N = \mathbf{I}_{|J|} \otimes 1_{\mathcal{S}} + \frac{Q_{I,J}^N R_{I,J}^N}{\sqrt{N}}. \tag{127}$$

Then left-multiply by $H_{I,J}^N$ on both sides and rearrange slightly to get the result. □

8.4. **More elaborate identities.** We specialize and combine the basic identities.

8.4.1. For N, $(I,J) \in \mathcal{I}_N^{(2)}$ and $j_1, j_2 \in J$, we have

$$R_I^N(j_1,j_2) - \delta_{j_1 j_2} H_{I\setminus J} = \mathbf{f}_{j_1}\mathbf{f}_J^* \frac{\Delta R_{I,J}^N}{\sqrt{N}}\mathbf{f}_J \mathbf{f}_{j_2}^*, \tag{128}$$

$$R_I^N(j_1,j_2) - \delta_{j_1 j_2} H_{I\setminus J} - \frac{H_{I\setminus J} Q_{I,J,j_1,j_2}^N H_{I\setminus J}}{\sqrt{N}} = \mathbf{f}_{j_1}\mathbf{f}_J^* \frac{\Delta^2 R_{I,J}^N}{N}\mathbf{f}_J \mathbf{f}_{j_2}^* \tag{129}$$

by merely rewriting (126) in the cases $k = 1$ and $k = 2$, respectively, at the level of individual matrix entries.

8.4.2. For N and $(I,J) \in \mathcal{I}_N^{(2)}$ we have

$$\begin{aligned} N(F_I^N - F_{I\setminus J}^N) &= \mathrm{tr}_{\mathcal{S}}(R_{I,J}^N) + \mathrm{tr}_{\mathcal{S}}\left(R_{I\setminus J}^N \frac{X}{\sqrt{N}}\mathbf{f}_J^* R_{I,J}^N \mathbf{f}_J \frac{X}{\sqrt{N}} R_{I\setminus J}^N\right) \\ &= \left(\mathrm{tr}_{\mathcal{S}} + T_{I\setminus J}^N \circ \Phi \circ \mathrm{tr}_{\mathcal{S}} + \frac{P_{I,J}^N}{\sqrt{N}}\right)(R_{I,J}^N) \end{aligned} \tag{130}$$

by applying $\mathrm{tr}_{\mathcal{S}}$ to both sides of (125). We note also the identity

$$H_I^N - H_{I\setminus J}^N = H_I^N \mathbf{1}_{\left[\left[E_{I\setminus J}^N\right]\right] \geq 1/2} - H_{I\setminus J}^N \mathbf{1}_{[\![E_I^N]\!] \geq 1/2} + H_I^N \Phi(F_I^N - F_{I\setminus J}^N) H_{I\setminus J}^N \tag{131}$$

obtained by exploiting the resolvent identity (35) in evident fashion.

8.4.3. For N and $I \in \mathcal{I}_N$ such that $|I| \geq 2$ we have

$$E_I^N + \frac{|I|-N}{N} 1_{\mathcal{S}} = \frac{1}{N}\sum_{j\in I}\left(\Phi(F_I^N - F_{I\setminus j}^N)R_{I,j}^N - \frac{Q_{I,j}^N R_{I,j}^N}{\sqrt{N}}\right) \tag{132}$$

after applying $\frac{1}{N}\sum_{j\in I}(\cdot)$ to both sides of (127) in the singleton case $J = \{j\}$ and rearranging.

Remark 8.4.4. Identity (132) is an approximate version of the Schwinger-Dyson equation. Identities of this sort have long been in use for study of Wigner matrices.

8.4.5. For N and $I \in \mathcal{I}_N$ such that $|I| \geq 2$ we also have

$$\begin{aligned} & E_I^N + \frac{|I|-N}{N} 1_{\mathcal{S}} + \frac{1}{N}\sum_{j\in I}\frac{Q_{I,j}^N H_{I\setminus j}^N}{\sqrt{N}} \\ = \;& \frac{1}{N}\sum_{j\in I}\left(\Phi(F_I^N - F_{I\setminus j}^N)R_{I,j}^N - \frac{Q_{I,j}^N \Delta R_{I,j}^N}{N}\right), \end{aligned} \tag{133}$$

by (126) for $k = 1$ in the singleton case $J = \{j\}$ and (132), after rearrangement.

8.4.6. For N and $I \in \mathcal{I}_N$ we have

$$H_I^N - F_I^N = H_I^N E_I^N - F_I^N \mathbf{1}_{[\![E_I^N]\!]\geq 1/2} \tag{134}$$

by direct appeal to the definitions. One then obtains for $|I| \geq 2$ the identity

$$\begin{aligned} & H_I^N - F_I^N + \frac{1}{N}\sum_{j\in I}\frac{F_{I\setminus j}^N Q_{I,j}^N H_{I\setminus j}^N}{\sqrt{N}} \\ = \;& \frac{N-|I|}{N} F_I^N + H_I^N (E_I^N)^2 - (F_I^N + F_I^N E_I^N)\mathbf{1}_{[\![E_I^N]\!]\geq 1/2} \\ & + \frac{1}{N}\sum_{j\in I}\left(F_I^N \Phi(F_I^N - F_{I\setminus j}^N)R_{I,j}^N - \frac{F_I^N Q_{I,j}^N \Delta R_{I,j}^N}{N} - \frac{(F_I^N - F_{I\setminus j}^N)Q_{I,j}^N H_{I\setminus j}^N}{\sqrt{N}}\right) \end{aligned} \tag{135}$$

by iterating (134) and combining it with (133).

8.4.7. For N and $(I, J) \in \mathcal{I}_N^{(2)}$ we have

$$\begin{aligned} & F_I^N - F_{I\setminus J}^N - |J|\frac{(1_{B(\mathcal{S})} + T_{I\setminus J}^N \circ \Phi)(H_{I\setminus J}^N)}{N} \\ = \;& \frac{P_{I,J}^N(R_{I,J}^N) + (\mathrm{tr}_{\mathcal{S}} + T_{I\setminus J}^N \circ \Phi \circ \mathrm{tr}_{\mathcal{S}})\left(\Delta R_{I,J}^N\right)}{N^{3/2}} \end{aligned} \tag{136}$$

by rearrangement of (130), using (126) for $k = 1$.

8.4.8. For N and $I \in \mathcal{I}_N$ we have

$$\begin{aligned} & (N+1)F_I^{N+1} - NF_I^N - \frac{1}{2}(F_I^N + T_I^N(\Lambda)) \\ = \;& \frac{1}{N}\mathrm{tr}_{\mathcal{S}}\Bigg(\left(N\delta_N - \frac{1}{2}\right)(\mathbf{e}_I + R_I^N(\mathbf{I}_\infty \otimes \Lambda))R_I^N \\ & \qquad + \frac{(N\delta_N)^2}{N}(\mathbf{e}_I + R_I^N(\mathbf{I}_\infty \otimes \Lambda))^2 R_I^{N+1}\Bigg) \end{aligned} \tag{137}$$

by Lemma 8.3.1 in the case $k = 2$ after using Remark 8.2.8, applying $\mathrm{tr}_{\mathcal{S}}$ on both sides and rearranging. Note that $\frac{1}{2} - \frac{1}{2N} \leq N\delta_N \leq \frac{1}{2}$.

8.4.9. Let

$$\mathrm{Link}^N = \frac{1}{2}(F_N^N + T_N^N(\Lambda)) - F_{N+1}^{N+1} + H_N^{N+1} + T_N^{N+1}(\Phi(H_N^{N+1})),$$

where here and below in similar contexts we abuse notation by writing N where we should more correctly write $\{1, \dots, N\}$. We then have

$$\begin{aligned} &N(F_{N+1}^{N+1} - F_N^N) - \mathrm{Link}^N \\ = \;& (N+1)F_N^{N+1} - NF_N^N - \frac{1}{2}(F_N^N + T_N^N(\Lambda)) \\ &+(N+1)(F_{N+1}^{N+1} - F_N^{N+1}) - H_N^{N+1} - T_N^{N+1}(\Phi(H_N^{N+1})) \end{aligned} \tag{138}$$

by mere rearrangement of terms.

8.5. **The bias identity.** We derive the most intricate identity used in the paper.

8.5.1. We first need an intermediate result which continues the process of expansion begun in identity (133). For N and $I \in \mathcal{I}_N$ such that $|I| \geq 2$, we have

$$\begin{aligned} &E_I^N + \frac{|I| - N}{N} 1_{\mathcal{S}} \\ &+\frac{1}{N}\sum_{j \in I}\left(\frac{(Q_{I,j}^N H_{I\setminus j}^N)^2 - (\Phi + \Phi \circ T_{I\setminus j}^N \circ \Phi)(R_{I,j}^N)R_{I,j}^N}{N} + \frac{(Q_{I,j}^N H_{I\setminus j}^N)^3}{N^{3/2}}\right) \\ = \;& \frac{1}{N}\sum_{j \in I}\left(-\frac{Q_{I,j}^N H_{I\setminus j}^N}{\sqrt{N}} - \frac{Q_{I,j}^N \Delta^3 R_{I,j}^N}{N^2} + \frac{\Phi \circ P_{I,j}^N(R_{I,j}^N)R_{I,j}^N}{N^{3/2}}\right) \end{aligned} \tag{139}$$

by expanding the terms $\frac{Q_{I,j}^N R_{I,j}^N}{\sqrt{N}}$ in (132) by using (126) for $k = 3$ in the singleton case $J = \{j\}$, and furthermore expanding the terms $\Phi(F_I^N - F_{I\setminus j}^N)R_{I,j}^N$ in (132) by using (130) in the singleton case $J = \{j\}$, after suitable rearrangement.

8.5.2. Fix $N \geq 2$ and $j \in N$ arbitrarily. To compactify notation put

$$\begin{aligned} \widetilde{T}_j^N &= \Phi + \Phi \circ T_{N\setminus j}^N \circ \Phi, \;\; \widetilde{P}_j^N \; = \; \Phi \circ P_{N\setminus j}^N, \;\; \check{R}_j^N = H_{N\setminus j}^N Q_{N,j}^N H_{N\setminus j}^N, \\ \mathrm{Err}_j^N &= (Q_{N,j}^N H_{N\setminus j}^N)^2 - \widetilde{T}_j^N(H_{N\setminus j}^N)H_{N\setminus j}^N + (Q_{N,j}^N H_{N,j}^N)^3/\sqrt{N}, \\ \mathrm{Err}_j^{N,1} &= (\widetilde{T}_j^N(H_{N\setminus j}^N)\check{R}_j^N + \widetilde{T}_j^N(\check{R}_j^N)H_{N\setminus j}^N + \widetilde{P}_j^N(H_{N\setminus j}^N)H_{N\setminus j}^N)/N - Q_{N,j}^N H_{N\setminus j}^N, \\ \mathrm{Err}_j^{N,2} &= \widetilde{T}_j^N(H_{N\setminus j}^N)\Delta^2 R_{N,j}^N + \widetilde{T}_j^N(\check{R}_j^N)\Delta R_{N,j}^N + \widetilde{T}_j^N(\Delta^2 R_{N,j}^N)R_{N,j}^N \\ &\quad +\widetilde{P}_j^N(H_{N\setminus j}^N)\Delta R_{N,j}^N + \widetilde{P}_j^N(\Delta R_{N,j}^N)R_{N,j}^N - Q_{N,j}^N \Delta^3 R_{N,j}^N. \end{aligned}$$

At last, we obtain the *bias identity*

$$E_N^N + \frac{1}{N}\sum_{j \in N}\frac{\mathrm{Err}_j^N}{N} = \frac{1}{N}\sum_{j \in N}\left(\frac{\mathrm{Err}_j^{N,1}}{\sqrt{N}} + \frac{\mathrm{Err}_j^{N,2}}{N^2}\right) \tag{140}$$

by using (126) several times with $k = 1, 2$ in the singleton case $J = \{j\}$ to expand the terms $(\Phi + \Phi \circ T_{N\setminus j}^N \circ \Phi)(R_{N,j}^N)R_{N,j}^N$ and $\Phi \circ P_{N,j}^N(R_{N,j}^N)R_{N,j}^N$ in (139), after suitable rearrangement.

9. L^p estimates for the block Wigner model

We introduce a straightforward generalization of the usual Wigner matrix model with matrix entries in a block algebra. Using all the tools collected in §4 and §8, we investigate how control of moments of "randomized resolvents" propagates to give control of moments of many related random variables. Our main result in this section is Theorem 9.4.2 below which converts identity (140) above to a key approximation.

9.1. **The block Wigner model.** The *ad hoc* infinite matrix formalism of §8.1 will be the algebraic framework for our discussion of the block Wigner model.

9.1.1. *Data.* Data for the block Wigner model consist of

- a block algebra $\mathcal{S}$,
- a random matrix $X \in \mathrm{Mat}_\infty(\mathcal{S})_{\mathrm{sa}}$,
- a (deterministic) linear map $\Phi \in B(\mathcal{S})$,
- a (deterministic) tensor $\Psi \in \mathcal{S}^{\otimes 2}$,
- a random element $\Lambda \in \mathcal{S}$ and
- a random variable $\mathfrak{G} \in [1,\infty)$.

9.1.2. *The σ-fields $\mathcal{F}(i,j)$ and the auxiliary random variables $\mathbf{z}$ and $\mathbf{t}$.* In addition to the data above, as a convenience of bookkeeping, we keep for use in the present setup the same system $\{\mathcal{F}(i,j)\}_{1\leq i\leq j<\infty}$ of independent σ-fields mentioned in §2.4.1. As before, let $\mathcal{F}$ denote the σ-field generated by all the $\mathcal{F}(i,j)$. More generally, for any set I of positive integers, let $\mathcal{F}_I$ denote the σ-field generated by the family $\{\mathcal{F}(i,j) \mid i,j \in I\}$. We also keep the random variables $\mathbf{z}$ from Theorem 2.4.6 and $\mathbf{t}$ from Remark 8.2.2 on hand, and we continue to assume that $\mathcal{F}$, $\mathbf{z}$ and $\mathbf{t}$ are independent. In §9 the random variables $\mathbf{z}$ and $\mathbf{t}$ intervene only through their σ-fields but in §10 these random variables themselves take an active part.

9.1.3. *Assumptions.* Of the sextuple $(\mathcal{S}, X, \Phi, \Psi, \Lambda, \mathfrak{G})$ we assume the following:

$$\sup_{i,j=1}^{\infty} \|[\![X(i,j)]\!]\|_p < \infty \text{ for } 1 \leq p < \infty. \tag{141}$$

$$X(i,j) \text{ is } \mathcal{F}(i\wedge j, i\vee j)\text{-measurable and of mean zero for all } i \text{ and } j. \tag{142}$$

$$\Phi = (\zeta \mapsto \mathbb{E}X(i,j)\zeta X(j,i)) \text{ and } \Psi = \mathbb{E}(X(i,j)^{\otimes 2}) \text{ for distinct } i \text{ and } j. \tag{143}$$

$$[\![\Lambda]\!]_p < \infty \text{ for } p \in [1,\infty). \tag{144}$$

$$\Lambda \text{ is } \sigma(\mathbf{z},\mathbf{t})\text{-measurable and } \mathfrak{G} \text{ is } \sigma(\mathbf{z})\text{-measurable.} \tag{145}$$

$$\mathbf{f}_I\left(\frac{X}{\sqrt{N}} - \mathbf{I}_\infty \otimes \Lambda\right)\mathbf{f}_I^* \in \mathrm{GL}_{|I|}(\mathcal{S}) \text{ for } N \text{ and } I \in \mathcal{I}_N. \tag{146}$$

For N and $I \in \mathcal{I}_N$ we then put

$$R_I^N = \mathbf{f}_I^*\left(\mathbf{f}_I\left(\frac{X}{\sqrt{N}} - \mathbf{I}_\infty \otimes \Lambda\right)\mathbf{f}_I^*\right)^{-1}\mathbf{f}_I \in \mathrm{Mat}_\infty(\mathcal{S}),$$

which is a generalized resolvent (Green's function). Finally, we assume that

$$\sup_N \bigvee_{I\in\mathcal{I}_N} \left\|[\![R_I^N/\mathfrak{G}]\!]\right\|_p < \infty. \tag{147}$$

We work with a fixed instance $(\mathcal{S}, X, \Phi, \Psi, \Lambda, \mathfrak{G})$ of the block Wigner model over $\mathcal{S}$ for the rest of §9.

Remark 9.1.4. Fix real numbers a and b. Using assumptions (14)—(20) it is easy to verify directly that the collection

$$\begin{aligned}(148)\qquad & (\mathcal{S}, X, \Phi, \Psi, \Lambda, \mathfrak{G}) \\ &= \left(\mathbb{C}, \bigcup_N (a\Xi_1^N + b\Xi_2^N), a^2+b^2, a^2-b^2, \mathbf{z}+\mathrm{i}\mathbf{t}, 1+1/\Im\mathbf{z}\right)\end{aligned}$$

satisfies assumptions (141)—(147). This specialization is merely a slight variant of the standard Wigner matrix model. The reader might find it helpful to concentrate on this case when making a first pass through the L^p estimates assembled below.

Remark 9.1.5. All instances of the block Wigner model needed for the proof of Theorem 2.4.6 arise as follows. Let $(\mathcal{S}, L, \Theta, e)$ be a SALT block design and let c_0, c_1, c_2 and $\mathfrak{T}$ be the constants from Definition 5.4.1. We keep the notation of Remark 8.2.2. Let Φ_L and Ψ_L be as in Definition 5.2.5. Then, using assumptions (14)—(20) along with Remark 5.4.6, it is easy to verify that the collection

$$\begin{aligned}(149)\qquad & (\mathcal{S}, X, \Phi, \Psi, \Lambda, \mathfrak{G}) \\ &= \left(\mathcal{S}, \bigcup_N L(\Xi^N), \Phi_L, \Psi_L, \Theta + \mathbf{z}e + \mathrm{i}\mathbf{t}1_{\mathcal{S}}, c_0(1+[\![L(\Xi)]\!])^{c_1}(1+1/\Im\mathbf{z})^{c_2}\right)\end{aligned}$$

satisfies assumptions (141)—(147).

9.1.6. *Random variables defined by recipes.* Since assumption (146) is a verbatim repetition of assumption (118), all the recipes of §8.2 define random variables in the present setting. The object R_I^N figuring in assumption (147) is of course a recipe. We now furthermore have random variables F_I^N, H_I^N, T_I^N, U_I^N, etc. at our disposal. The compound objects Link^N, Err_j^N, etc. figuring in the more elaborate identities also become random variables in the present setting.

Remark 9.1.7. We note that (147) can be considerably refined in the specialization (149) of the block Wigner model. Namely, for each N and $I \in \mathcal{I}_N$ we have almost sure bounds

$$[\![F_I^N]\!] \leq [\![R_I^N]\!] \leq \mathfrak{G}\left(1 + \left[\!\!\left[\frac{\mathbf{f}_I X \mathbf{f}_I^*}{\sqrt{N}}\right]\!\!\right]\right)^{c_1}, \tag{150}$$

$$[\![F_I^N]\!] \leq [\![R_I^N]\!] \leq \frac{1}{2} \text{ on the event } \mathbf{t} \geq \mathfrak{T}. \tag{151}$$

One can also easily verify that

$$(G_L : \mathcal{D}_L \to \mathcal{S}, \Phi_L, \Lambda, \mathfrak{T}, \mathfrak{G}) \text{ is an SD tunnel.} \tag{152}$$

Of course this SD tunnel is random since Λ and $\mathfrak{G}$ are random.

9.1.8. *Partially averaged random variables.* We complete our enumeration of the random variables we will be studying. For N and $I \in \mathcal{I}_N$ we define

$$\overline{F}_I^N = \mathfrak{G}\mathbb{E}(F_I^N/\mathfrak{G}|\mathbf{z}, \mathbf{t}) \in \mathcal{S} \text{ and } \overline{E}_I^N = 1_{\mathcal{S}} + (\Lambda + \Phi(\overline{F}_I^N))\overline{F}_I^N \in \mathcal{S}.$$

Since $[\![F_I^N]\!]/\mathfrak{G}$ is integrable by assumption (147), in fact $\overline{F}_I^N$ and $\overline{E}_I^N$ are well-defined, almost surely. Theorem 9.4.2 below gives a delicate approximation to $\overline{E}_N^N$. (Recall our abuse of notation $N = \{1, \ldots, N\}$.)

9.2. **Basic estimates.** We start gathering consequences of (147).

9.2.1. *The norms* $\|\!|\cdot|\!\|_{p,k}$. Given a constant $p \in [1,\infty)$, a positive integer k and a finite-dimensional-Banach-space-valued random variable Z (defined on the same probability space as $\mathfrak{G}$), we write $\|\!|Z|\!\|_{p,k} = \left\|[\![Z]\!]/\mathfrak{G}^k\right\|_p$ to compress notation.

Remark 9.2.2. We emphasize that in §9 we make no assumption concerning the strength of the repulsion of $\mathbf{z}$ from the real axis. Indeed, we make no assumptions about $\mathbf{z}$ at all. But nevertheless, looking ahead to the completion of the proof of Theorem 2.4.6, we are obliged to keep track of issues involving the repulsion strength. To do so, we will use the parameter k appearing in the norm $\|\!|\cdot|\!\|_{p,k}$ and in a similar seminorm introduced in §9.2.4 below.

Proposition 9.2.3. *For each constant* $p \in [1,\infty)$ *we have*

$$\sup_N \bigvee_{I \in \mathcal{I}_N} \|\!|R_I^N|\!\|_{p,1} \vee \|\!|F_I^N|\!\|_{p,1} \vee \|\!|H_I^N|\!\|_{p,1} \vee \|\!|T_I^N|\!\|_{p,2} \vee \|\!|U_I^N|\!\|_{p,2} < \infty, \tag{153}$$

$$\sup_N \bigvee_{(I,J)\in\mathcal{I}_N^{(2)}} \|\!|R_{I,J}^N|\!\|_{p,1} \vee \|\!|H_{I,J}^N|\!\|_{p,1} < \infty, \tag{154}$$

$$\sup_N \bigvee_{(I,J)\in\mathcal{I}_N^{(2)}} N\left\|\!\left|F_I^N - F_{I\setminus J}^N\right|\!\right\|_{p,3} < \infty. \tag{155}$$

Proof. The claim made in (153) for R_I^N just repeats the hypothesis (147) in different notation. We have

$$[\![R_I^N]\!] \geq [\![F_I^N]\!] \vee [\![T_I^N]\!]^{1/2} \vee \frac{1}{2}[\![H_I^N]\!] \vee \frac{1}{\sqrt{s}}[\![U_I^N]\!]^{1/2}$$

obviously in the first two cases, by Lemma 3.1.1 in the penultimate case and Lemma 4.2.3 in the last, where s^2 is the dimension of $\mathcal{S}$ over the complex numbers. Thus (153) holds in general. Clearly, we have

$$[\![R_{I,J}^N]\!] \leq [\![R_I^N]\!] \quad \text{and} \quad [\![H_{I,J}^N]\!] = \left[\!\!\left[H_{I\setminus J}^N\right]\!\!\right],$$

whence (154) via (153). By Lemma 4.2.3 and identity (130) we have

$$\frac{N}{|J|^2}\left[\!\!\left[F_I^N - F_{I\setminus J}^N\right]\!\!\right] \leq [\![R_I^N]\!] + \frac{s}{N}[\![R_I^N]\!]\left[\!\!\left[R_{I\setminus J}^N\right]\!\!\right]^2 \sum_{(i,j)\in(I\setminus J)\times J} [\![X(i,j)]\!]^2.$$

From this, estimate (155) follows by assumption (141), the Minkowski inequality and estimate (153). $\square$

9.2.4. *The seminorms* $\|\!|\cdot|\!\|_{p,k,I}$. Given a constant $p \in [1,\infty)$, a positive integer k, a set I of positive integers and a finite-dimensional-Banach-space-valued random variable Z defined on the same probability space as $\mathfrak{G}$ such that $\|\!|Z|\!\|_{p,k} < \infty$, we define

$$\|\!|Z|\!\|_{p,k,I} = \left\|\,[\![\mathbb{E}(Z/\mathfrak{G}^k|\mathcal{F}_I,\mathbf{z},\mathbf{t})]\!]\,\right\|_p.$$

Since the random variable $[\![Z/\mathfrak{G}^k]\!]$ is assumed to be in $L^p \subset L^1$, the conditional expectation appearing on the right is well-defined, almost surely, and moreover

$$\|\!|Z|\!\|_{p,k} \geq \|\!|Z|\!\|_{p,k,I} \geq \|\!|Z|\!\|_{p,k,J}$$

for any set $J \subset I$ by Jensen's inequality. In particular,

$$\left\|\mathbb{E}(Z/\mathfrak{G}^k|\mathbf{z},\mathbf{t})\right\|_p = \|\!|Z|\!\|_{p,k,\emptyset}$$

whenever $\|\!|Z|\!\|_{p,k} < \infty$.

Proposition 9.2.5. *For each constant $p \in [1, \infty)$ we have*

$$\sup_N \bigvee_{(I,J)\in\mathcal{I}_N^{(2)}} \left\|\left|Q_{I,J}^N\right|\right\|_{p,1} \vee \left\|\left|P_{I,J}^N\right|\right\|_{p,2} < \infty, \tag{156}$$

$$\sup_N \bigvee_{(I,J)\in\mathcal{I}_N^{(2)}} \left\|\left|Q_{I,J}^N\right|\right\|_{p,1,I\setminus J} \vee \left\|\left|P_{I,J}^N\right|\right\|_{p,2,I\setminus J} = 0. \tag{157}$$

It is hard to overestimate the importance of this proposition. This is the estimate ultimately driving convergence. Our exploitation of it is of course an imitation of the procedure of [3].

Proof. Fix N, $(I, J) \in \mathcal{I}_N^{(2)}$ and $j_1, j_2 \in J$ arbitrarily. By definition we have

$$Q_{I,J,j_1,j_2}^N = -X(j_1, j_2) + \frac{1}{\sqrt{N}}\left(\mathbf{f}_{j_1} X R_{I\setminus J}^N X \mathbf{f}_{j_2}^* - N\delta_{j_1 j_2}\Phi(F_{I\setminus J}^N)\right), \tag{158}$$

$$P_{I,J,j_1,j_2}^N = \left(\zeta \mapsto \frac{1}{\sqrt{N}}\left(\mathrm{tr}_{\mathcal{S}}(R_{I\setminus J}^N X \mathbf{f}_{j_1}^* \zeta \mathbf{f}_{j_2} X R_{I\setminus J}^N) - N\delta_{j_1 j_2} T_{I\setminus J}^N(\Phi(\zeta))\right)\right).$$

By (141) and (153), the random variables $\left[\left[Q_{I,J,j_1,j_2}^N\right]\right]$ and $\left[\left[P_{I,J,j_1,j_2}^N\right]\right]$ are integrable, hence the conditional expectations

$$\mathbb{E}(Q_{I,J,j_1,j_2}^N/\mathfrak{G}|\mathcal{F}_{I\setminus J}, \mathbf{z}, \mathbf{t}) \text{ and } \mathbb{E}(P_{I,J,j_1,j_2}^N/\mathfrak{G}^2|\mathcal{F}_{I\setminus J}, \mathbf{z}, \mathbf{t})$$

are well-defined and vanish almost surely by assumptions (141), (142) and (143). By Proposition 4.2.2, Remark 4.2.6, estimate (153) and the hypotheses of the block Wigner model, the quantities

$$\left\|\left|Q_{I,J,j_1,j_2}^N + X(j_1, j_2)\right|\right\|_{p,1}, \quad \left\|\left|X(j_1, j_2)\right|\right\|_{p,1} \text{ and } \left\|\left|P_{I,J,j_1,j_2}^N\right|\right\|_{p,2}$$

are bounded uniformly in N, I, J, j_1 and j_2. Thus claims (156) and (157) hold. □

9.3. More elaborate estimates.

We combine and specialize the basic estimates.

Proposition 9.3.1. *For each constant $p \in [1, \infty)$ we have*

$$\sup_N \bigvee_{\substack{I\in\mathcal{I}_N \\ \text{s.t. } |I|\geq 2}} \sqrt{N}\left\|\left|E_I^N + \frac{|I| - N}{N} 1_{\mathcal{S}}\right|\right\|_{p,4} < \infty, \tag{159}$$

$$\sup_N \bigvee_{\substack{(I,J)\in\mathcal{I}_N^{(2)} \text{ s.t.} \\ |I|\geq N-\sqrt{N}}} \left\|\left|\Delta_{I,J}^N\right|\right\|_{p,4} < \infty, \tag{160}$$

$$\sup_N \bigvee_{\substack{I\in\mathcal{I}_N \text{ s.t.} \\ |I|\geq N-99}} N\left\|\left|E_I^N\right|\right\|_{p,6,\emptyset} < \infty. \tag{161}$$

Proof. We take Propositions 9.2.3 and 9.2.5 for granted at every step. Identity (132) implies the estimate (159). Estimate (159) and the Chebychev bound

$$\mathbf{1}_{\left[\left[E_I^N\right]\right]\geq 1/2} \leq (2\left[\left[E_I^N\right]\right])^c \quad (c \geq 0) \tag{162}$$

imply estimate (160). Identity (133) and estimate (160) imply the estimate

$$\sup_N \bigvee_{\substack{I\in\mathcal{I}_N \\ \text{s.t. } |I|\geq N-\sqrt{N}}} N\left\|\left|E_I^N + \frac{|I| - N}{N} 1_{\mathcal{S}} + \frac{1}{N}\sum_{j\in I}\frac{Q_{I,j}^N H_{I\setminus j}^N}{\sqrt{N}}\right|\right\|_{p,6} < \infty.$$

Estimate (161) follows via (157). □

Proposition 9.3.2. *For each constant $p \in [1,\infty)$ we have*

$$\sup_N N^{3/2}\left\|\left\|F_{N+1}^{N+1} - F_N^N - \frac{\mathrm{Link}^N}{N}\right\|\right\|_{p,7} < \infty, \tag{163}$$

$$\sup_N \bigvee_{\substack{I\in\mathcal{I}_N \text{ s.t.}\\ |I|\geq N-99}} N^2\left\|\left\|\overline{E}_I^N - E_I^N\right\|\right\|_{p,14,\emptyset} < \infty. \tag{164}$$

Proof. We take Propositions 9.2.3, 9.2.5 and 9.3.1 for granted at every step. We have

$$\sup_N N\left\|\left\|(N+1)F_N^{N+1} - NF_N^N - \frac{1}{2}(F_N^N + T_N^N(\Lambda))\right\|\right\|_{p,3} < \infty$$

by identity (137) along with assumption (144). The estimate

$$\sup_N \bigvee_{\substack{(I,J)\in\mathcal{I}_N^{(2)}\\ |I|\geq N-\sqrt{N}}} N^{3/2}\left\|\left\|F_I^N - F_{I\setminus J}^N - |J|\frac{(1_{B(\mathcal{S})} + T_{I\setminus J}^N\circ\Phi)(H_{I\setminus J}^N)}{N}\right\|\right\|_{p,7} < \infty$$

follows from identity (136). Estimate (163) then follows via the definition of Link^N. From the last estimate above it also follows that

$$\sup_N \sup_{\substack{I\in\mathcal{I}_N\\ |I|\geq N-99}} N^2\left\|\mathbf{Var}_{\mathcal{S}}(F_I^N/\mathfrak{G}^7|\mathbf{z},\mathbf{t})\right\|_p < \infty$$

via Proposition 4.3.2, whence estimate (164). □

Proposition 9.3.3. *For each constant $p \in [1,\infty)$ we have*

$$\sup_N \bigvee_{\substack{(I,J)\in\mathcal{I}_N^{(2)} \text{ s.t.}\\ |I|\geq N-\sqrt{N}}} \bigvee_{j_1,j_2\in J} \sqrt{N}\left\|\left\|R_I^N(j_1,j_2) - \delta_{j_1 j_2}H_{I\setminus J}^N\right\|\right\|_{p,5} < \infty, \tag{165}$$

$$\sup_N \bigvee_{\substack{(I,J)\in\mathcal{I}_N^{(2)} \text{ s.t.}\\ |I|\geq N-\sqrt{N}}} \bigvee_{j_1,j_2\in J} N\left\|\left\|R_I^N(j_1,j_2) - \delta_{j_1,j_2}H_{I\setminus J}^N\right\|\right\|_{p,9,I\setminus J} < \infty, \tag{166}$$

$$\sup_N \bigvee_{\substack{(I,J)\in\mathcal{I}_N^{(2)} \text{ s.t.}\\ |I|\geq N-\sqrt{N}}} N\left\|\left\|H_I^N - H_{I\setminus J}^N\right\|\right\|_{p,9} < \infty, \tag{167}$$

$$\sup_N \bigvee_{\substack{I\in\mathcal{I}_N \text{ s.t.}\\ |I|\geq N-\sqrt{N}}} \sqrt{N}\left\|\left\|H_I^N - F_I^N\right\|\right\|_{p,5} < \infty, \tag{168}$$

$$\sup_N \bigvee_{\substack{I\in\mathcal{I}_N\\ |I|\geq N-\sqrt{N}}} N\left\|\left\|H_I^N - F_I^N\right\|\right\|_{p,9,I\setminus J} < \infty. \tag{169}$$

Proof. Taking Propositions 9.2.3, 9.2.5 and 9.3.1 for granted and using again the Chebychev bound (162), one derives the estimates in question from identities (128), (129), (131), (134) and (135), respectively. □

9.4. **The bias theorem.** We work out a delicate approximation to $\overline{E}_N^N$. We use again the apparatus introduced to state and prove Proposition 4.4.5, as well as the cumulant and shuffle notation introduced in §6.5.

9.4.1. *Corrections.* For $N \geq 2$ and $j = 1, \ldots, N$ we define

$$\begin{aligned}\mathrm{Corr}_j^N &= \left\langle [\Psi, \Psi]_2, [U^N_{N\setminus j}, (H^N_{N\setminus j})^{\otimes 2}]_2 \right\rangle_4 - \Phi(H^N_{N\setminus j})H^N_{N\setminus j}\\ &\quad + \left\langle \mathbb{E}X(j,j)^{\otimes 2}, (H^N_{N\setminus j})^{\otimes 2}\right\rangle_2 - \frac{1}{\sqrt{N}}\left\langle \mathbb{E}X(j,j)^{\otimes 3}, (H^N_{N\setminus j})^{\otimes 3}\right\rangle_3\\ &\quad + \frac{1}{N}\sum_{i \in N\setminus j}\left\langle \mathbf{C}^{(4)}(X(i,j)), [(R^N_{I\setminus j,i})^{\otimes 2}, (H^N_{N\setminus j})^{\otimes 2}]_2\right\rangle_4.\end{aligned}$$

Theorem 9.4.2. *For each constant $p \in [1,\infty)$ we have*

$$\sup_{N\geq 2} N^2 \left\|\left| \overline{E}_N^N + \frac{1}{N}\sum_{j=1}^N \frac{\mathrm{Corr}_j^N}{N}\right|\right\|_{p,14,\emptyset} < \infty. \tag{170}$$

The proof of the theorem takes up the rest of §9.4. We need several lemmas.

Lemma 9.4.3. *For each constant $p \in [1,\infty)$ we have*

$$\sup_{N\geq 2}\bigvee_{j=1}^N \left\|\left|\mathrm{Err}_j^N\right|\right\|_{p,6} \vee \left\|\left|\mathrm{Err}_j^{N,1}\right|\right\|_{p,6} \vee \left\|\left|\mathrm{Err}_j^{N,2}\right|\right\|_{p,14} < \infty, \tag{171}$$

$$\sup_{N\geq 2}\bigvee_{j=1}^N \left\|\left|\mathrm{Err}_j^{N,1}\right|\right\|_{p,6,N\setminus j} = 0. \tag{172}$$

Proof. Taking Propositions 9.2.3, 9.2.5 and 9.3.1 for granted, these facts can be read off from the definitions presented in §8.5. □

9.4.4. *Moment notation.* For any sequence $\mathbf{i} = i_1 \cdots i_{2k}$ of positive integers and positive integer j not appearing in $\mathbf{i}$ put

$$\begin{aligned}\mathbf{M}_j(\mathbf{i}) &= \mathbb{E}\Big[(X(j,i_1)\otimes X(i_2,j) - \mathbb{E}(X(j,i_1)\otimes X(i_2,j)))\otimes\cdots\\ &\quad \cdots\otimes(X(j,i_{2k-1})\otimes X(i_{2k},j) - \mathbb{E}(X(j,i_{2k-1})\otimes X(i_{2k},j)))\Big] \in \mathcal{S}^{\otimes 2k}.\end{aligned}$$

Lemma 9.4.5. *For sequences $\mathbf{i} = i_1\cdots i_{2k}$ of positive integers, and positive integers j not appearing in $\mathbf{i}$, the following statements hold:*

(I) *For each fixed k, $[\![\mathbf{M}_j(\mathbf{i})]\!]$ is bounded uniformly in $\mathbf{i}$ and j.*
(II) *$\mathbf{M}_j(\mathbf{i})$ vanishes unless $\Pi(\mathbf{i}) \in \mathrm{Part}^*(2k)$.*
(III) *If $\Pi(\mathbf{i}) \in \mathrm{Part}_2^*(2k)$, then $\mathbf{M}_j(\mathbf{i})$ depends only on $\Pi(\mathbf{i})$.*

Proof. Assumption (141) implies statement (I). Assumptions (142) implies statement (II). Assumptions (142) and (143) imply statement (III). □

9.4.6. *Tensor products of resolvent entries.* For N, $I \in \mathcal{I}_N$ and sequences $\mathbf{i} = i_1\cdots i_{2k} \in \mathrm{Seq}(2k, I)$ put

$$R_I^N(\mathbf{i}) = R_I^N(i_1,i_2)\otimes\cdots\otimes R_I^N(i_{2k-1},i_{2k}) \in \mathcal{S}^{\otimes k}.$$

9.4.7. *The random variable* Rub_j^N. For $N \geq 2$ and $j = 1, \ldots, N$ put

$$\mathrm{Rub}_j^N = \frac{1}{N^2} \sum_{\substack{\mathbf{i} \in \mathrm{Seq}(6, N \setminus j) \text{ s.t.} \\ \Pi(\mathbf{i}) \in \mathrm{Part}^*(6) \text{ and} \\ \Pi(\mathbf{i}) \sim \{\{1,2,3\},\{4,5,6\}\}}} \langle \mathbf{M}_j(\mathbf{i}), [R^N_{N\setminus j}(\mathbf{i}), (H^N_{N\setminus j})^{\otimes 3}]_3 \rangle_6.$$

Here we employ again the notation $\sim$ for Γ_3-orbit equivalence previously introduced in connection with the list (55).

Lemma 9.4.8. *For $N \geq 3$ and $j = 1, \ldots, N$ we have*

$$\left[\left[\mathfrak{G}^6 \mathbb{E}(\mathrm{Err}_j^N / \mathfrak{G}^6 | \mathcal{F}_{N\setminus j}, \mathbf{z}, \mathbf{t}) - \mathrm{Corr}_j^N - \mathrm{Rub}_j^N\right]\right] \leq \frac{c}{N} \left[\left[R^N_{N\setminus j}\right]\right]^6,$$

almost surely, for a constant c independent of N and j.

Proof. In the case $(I, J) = (N, \{j\})$, formula (158) above simplifies to

$$Q^N_{N,j} + X(j,j) = \frac{1}{\sqrt{N}} \left(\mathbf{f}_j X R^N_{N\setminus j} X \mathbf{f}_j^* - \mathfrak{G}\mathbb{E}\left(\mathbf{f}_j X R^N_{N\setminus j} X \mathbf{f}_j^* / \mathfrak{G} | \mathcal{F}_{N\setminus j}, \mathbf{z}, \mathbf{t} \right) \right).$$

Note that the right side is independent of $X(j,j)$. A straightforward calculation using Lemma 9.4.5(II) yields that for $k \in \{2, 3\}$,

$$\begin{aligned} & \mathfrak{G}^{2k} \mathbb{E}((Q^N_{N,j} H^N_{N\setminus j})^k / \mathfrak{G}^{2k} | \mathcal{F}_{N\setminus j}, \mathbf{z}, \mathbf{t}) - (-1)^k \langle \mathbb{E} X(j,j)^{\otimes k}, (H^N_{N\setminus j})^{\otimes 3} \rangle_3 \\ = \ & \frac{1}{N^{k/2}} \sum_{\substack{\mathbf{i} \in \mathrm{Seq}(2k, N\setminus j) \\ \text{s.t. } \Pi(\mathbf{i}) \in \mathrm{Part}^*(2k)}} \langle \mathbf{M}_j(\mathbf{i}), [R^N_{N\setminus j}(\mathbf{i}), (H^N_{N\setminus j})^{\otimes k}]_k \rangle_{2k}. \end{aligned}$$

By a calculation using Lemma 9.4.5(II,III) and enumeration (54), with $\alpha, \beta \in N \setminus j$ arbitrarily chosen distinct elements, we have

$$\begin{aligned} & \mathfrak{G}^4 \mathbb{E}((Q^N_{N,j} H^N_{N\setminus j})^2 / \mathfrak{G}^4 | \mathcal{F}_{N\setminus j}, \mathbf{z}, \mathbf{t}) - \langle \mathbb{E}(X(j,j)^{\otimes 2}), (H^N_{N\setminus j})^{\otimes 2} \rangle_2 \\ = \ & \frac{1}{N} \sum_{i_1, i_2 \in N\setminus j} \langle \mathbf{M}_j(\alpha\beta\alpha\beta), [R^N_{N\setminus j}(i_1, i_2)^{\otimes 2}, (H^N_{N\setminus j})^{\otimes 2}]_2 \rangle_4 \\ & + \frac{1}{N} \sum_{i_1, i_2 \in N\setminus j} \langle \mathbf{M}_j(\alpha\beta\beta\alpha), [R^N_{N\setminus j}(i_1, i_2) \otimes R^N_{N\setminus j}(i_2, i_1), (H^N_{N\setminus j})^{\otimes 2}]_2 \rangle_4 \\ & + \frac{1}{N} \sum_{i \in N\setminus j} \left\langle \mathbf{M}_j(iiii) - \mathbf{M}_j(\alpha\beta\beta\alpha) - \mathbf{M}_j(\alpha\beta\alpha\beta), [(R^N_{N\setminus j,i})^{\otimes 2}, (H^N_{N\setminus j})^{\otimes 2}]_2 \right\rangle_4 \\ = \ & \left\langle [\Psi, \Psi]_2, [U^N_{N\setminus j}, (H^N_{N\setminus j})^{\otimes 2}]_2 \right\rangle_4 + \Phi \circ T^N_{N\setminus j} \circ \Phi(H^N_{N\setminus j}) \\ & + \frac{1}{N} \sum_{i \in N\setminus j} \langle \mathbf{C}^{(4)}(X(i,j)), [(R^N_{N\setminus j,i})^{\otimes 2}, (H^N_{N\setminus j})^{\otimes 2}]_2 \rangle_4. \end{aligned}$$

It follows that

$$\begin{aligned} & \mathfrak{G}^6 \mathbb{E}(\mathrm{Err}_j^N / \mathfrak{G}^6 | \mathcal{F}_{N\setminus j}, \mathbf{z}, \mathbf{t}) - \mathrm{Corr}_j^N - \mathrm{Rub}_j^N \\ = \ & \frac{1}{N^2} \sum_{\substack{\mathbf{i} \in \mathrm{Seq}(6, N\setminus j) \text{ s.t.} \\ \Pi(\mathbf{i}) \in \mathrm{Part}^*(6) \text{ and} \\ \Pi(\mathbf{i}) \not\sim \{\{1,2,3\},\{4,5,6\}\}}} \langle \mathbf{M}_j(\mathbf{i}), [R^N_{N\setminus j}(\mathbf{i}), (H^N_{N\setminus j})^{\otimes 3}]_3 \rangle_6, \end{aligned}$$

whence the result by Proposition 4.4.5 and Lemma 9.4.5(I,III). $\square$

Lemma 9.4.9. *Fix $p \in [1,\infty)$ arbitrarily. For $N \geq 3$ and distinct $j, j_1, j_2 \in N$, the quantity*

$$N\left|\left|\left|(H^N_{N\setminus j})^{\otimes 3} \otimes R^N_{N\setminus j}(j_1j_1j_2j_2j_1j_2)\right|\right|\right|_{p,14,N\setminus\{j,j_1,j_2\}}$$

is bounded uniformly in N, j, j_1 and j_2.

Proof. Put $J = \{j, j_1, j_2\}$. The quantity

$$N\left|\left|\left|(H^N_{N\setminus J})^{\otimes 5} \otimes R^N_{N\setminus j}(j_1, j_2)\right|\right|\right|_{p,14,N\setminus J}$$

is bounded uniformly in N, j, j_1 and j_2 by (153) and (166). The quantity

$$N\left|\left|\left|(H^N_{N\setminus j})^{\otimes 3} \otimes R^N_{N\setminus j}(j_1j_1j_2j_2j_1j_2) - (H^N_{N\setminus J})^{\otimes 5} \otimes R^N_{N\setminus j}(j_1, j_2)\right|\right|\right|_{p,14}$$

is bounded uniformly in N, j, j_1 and j_2 by (165) and (167). □

9.4.10. *Completion of the proof of Theorem 9.4.2.* We have

$$\sup_{N\geq 2} N^2\left|\left|\left|\overline{E}^N_N - E^N_N\right|\right|\right|_{p,14,\emptyset} < \infty$$

by estimate (164). We have

$$\sup_{N\geq 2} N^2\left|\left|\left|E^N_N + \frac{1}{N}\sum_{j=1}^N \frac{\mathrm{Err}^N_j}{N}\right|\right|\right|_{p,14,\emptyset} < \infty$$

by the bias identity and Lemma 9.4.3. We have

$$\sup_{N\geq 2} \bigvee_{j=1}^N N\left|\left|\left|\mathrm{Err}^N_j - \mathrm{Corr}^N_j - \mathrm{Rub}^N_j\right|\right|\right|_{p,14} < \infty$$

by Proposition 9.2.3 and Lemma 9.4.8. Finally, we have

$$\sup_{N\geq 2} \bigvee_{j=1}^N N\left|\left|\left|\mathrm{Rub}^N_j\right|\right|\right|_{p,14,\emptyset} < \infty$$

by Lemma 9.4.5(I) and Lemma 9.4.9, which finishes the proof. □

10. Concluding arguments

We finish the proof of Theorem 2.4.6.

10.1. **Setup for the concluding arguments.** Throughout §10 we fix an instance

$$(\mathcal{S}, L, \Theta, e), c_0, c_1, c_2, \mathfrak{T} \tag{173}$$

of Definition 5.4.1 and we work with the corresponding instance

$$\begin{aligned}&(\mathcal{S}, X, \Phi, \Psi, \Lambda, \mathfrak{G})\\ =\ &\left(\mathcal{S}, \bigcup_N L(\Xi^N), \Phi_L, \Psi_L, \Theta + \mathbf{z}e + \mathrm{i}\mathbf{t}1_{\mathcal{S}}, c_0(1 + [\![L(\Xi)]\!])^{c_1}(1 + 1/\Im\mathbf{z})^{c_2}\right)\end{aligned}$$

of the block Wigner model exhibited in Remark 9.1.5. We emphasize that we must consider a general example of a SALT block design because, at various stages below, we have to consider both a SALT block design arising as a self-adjoint linearization, i.e., via Proposition 5.4.2, and also a SALT block design arising by the underline construction, i.e., via Lemma 5.5.4.

Given any $\sigma(\mathcal{F},\mathbf{z},\mathbf{t})$-measurable Banach-space-valued integrable random variable Z, we define $Z|_{\mathbf{t}=0}$ to be any $\sigma(\mathcal{F},\mathbf{z})$-measurable random variable which on the event $\mathbf{t}=0$ equals Z almost surely. For example, we have $\Lambda|_{\mathbf{t}=0}=\Theta+\mathbf{z}e$. Since the latter random variable intervenes frequently below, we will write $\Lambda_0=\Lambda|_{\mathbf{t}=0}$ to compress notation.

In a similar vein, given Z as above along with a σ-field $\mathcal{G}$ on which $\mathbf{z}$ is measurable, we abuse notation by writing $\mathbb{E}(Z|\mathcal{G})=\mathfrak{G}^k\mathbb{E}(Z/\mathfrak{G}^k|\mathcal{G})$ when there exists some positive integer k large enough so that $Z/\mathfrak{G}^k$ is integrable and hence the conditional expectation $\mathbb{E}(Z/\mathfrak{G}^k|\mathcal{G})$ is well-defined.

We will employ the abbreviated notation

$$G:\mathcal{D}\to\mathcal{S}$$

in place of the more heavily subscripted notation $G_L:\mathcal{D}_L\to\mathcal{S}$. In a similar spirit we write

$$G'=\mathbf{D}[G],\ \ \check{G}=((G^{-1})^{\otimes 2}-\Psi)^{-1},\ \ \mathrm{Bias}^N=\mathrm{Bias}^N_L.$$

Note also that for every $p\in[1,\infty)$ the bounds

$$\|G(\Lambda)\|_{p,1}\vee\|G'(\Lambda)\|_{p,2}\vee\left\|\check{G}(\Lambda)\right\|_{p,2}\vee\sup_N\left\|\mathrm{Bias}^N(\Lambda)\right\|_{p,5}<\infty \tag{174}$$

hold, as one checks by means of Remark 7.1.3. We also have

$$\left\|G(\Lambda)^{-1}\right\|_{p,1}<\infty \tag{175}$$

by the SD equation $1_{\mathcal{S}}+(\Lambda+\Phi(G(\Lambda)))G(\Lambda)=0$ and assumption (144).

10.2. **Application of Proposition 7.3.2.** We claim that

$$\sup_N\bigvee_{\substack{I\in\mathcal{I}_N\text{ s.t.}\\|I|\geq N-99}}\sqrt{N}\left\|F_I^N|_{\mathbf{t}=0}-G(\Lambda_0)\right\|_{p,99}<\infty, \tag{176}$$

$$\sup_N\bigvee_{\substack{I\in\mathcal{I}_N\text{ s.t.}\\|I|\geq N-99}}N\left\|F_I^N|_{\mathbf{t}=0}+G'(\Lambda_0;(E_I^N|_{\mathbf{t}=0})G(\Lambda_0)^{-1})-G(\Lambda_0)\right\|_{p,99}<\infty, \tag{177}$$

$$\sup_N N^2\left\|\overline{F}_N^N|_{\mathbf{t}=0}+G'(\Lambda_0;(\overline{E}_N^N|_{\mathbf{t}=0})G(\Lambda_0)^{-1})-G(\Lambda_0)\right\|_{p,99}<\infty. \tag{178}$$

To prove the claim we introduce several further random variables. Put

$$\mathfrak{C}=99e^{2\mathfrak{T}}(1+[\![\Phi]\!]+[\![\Theta]\!]+|\mathbf{z}|).$$

For N and $I\in\mathcal{I}_N$ put

$$\mathfrak{L}_I^N=\mathfrak{G}^2\left(1+\left[\left[\frac{\mathbf{f}_IX\mathbf{f}_I^*}{\sqrt{N}}\right]\right]\right)^{2c_1},\ \ \mathfrak{E}_I^N=\mathbb{E}(\left[\left[E_I^N\right]\right]|\mathcal{F},\mathbf{z}).$$

Also for N put

$$\overline{\mathfrak{L}}^N=\mathfrak{G}^2\mathbb{E}\left(1+\left[\left[\frac{\mathbf{f}_NX\mathbf{f}_N^*}{\sqrt{N}}\right]\right]\right)^{2c_1},\ \ \overline{\mathfrak{E}}^N=\mathbb{E}\left(\left[\left[\overline{E}_N^N\right]\right]\middle|\mathbf{z}\right).$$

By Proposition 7.3.2 applied conditionally, with help from Remarks 5.4.6 and 9.1.7 to check hypotheses, we have

$$\begin{aligned}
&\sqrt{N}\left[\left[F_I^N|_{\mathbf{t}=0} - G(\Lambda_0)\right]\right] \leq \sqrt{N}(\mathfrak{C}\mathfrak{G}\mathfrak{L}_I^N)^6(\mathfrak{E}_I^N + (\mathfrak{E}_I^N)^2),\\
&N\left[\left[F_I^N|_{\mathbf{t}=0} + G'(\Lambda_0; (E_I^N|_{\mathbf{t}=0})G(\Lambda_0)^{-1}) - G(\Lambda_0)\right]\right]\\
&\leq N(\mathfrak{C}\mathfrak{G}\mathfrak{L}_I^N)^{12}((\mathfrak{E}_I^N)^2 + (\mathfrak{E}_I^N)^4),\\
&N^2\left[\left[\overline{F}_N^N|_{\mathbf{t}=0} + G'(\Lambda_0; (\overline{E}_N^N|_{\mathbf{t}=0})G(\Lambda_0)^{-1}) - G(\Lambda_0)\right]\right]\\
&\leq N^2(\mathfrak{C}\mathfrak{G}\overline{\mathfrak{L}}^N)^{12}((\overline{\mathfrak{E}}^N)^2 + (\overline{\mathfrak{E}}^N)^4).
\end{aligned}$$

Now fix $p \in [1,\infty)$ arbitrarily. The right sides above can be bounded in the norm $|\!|\!|\cdot|\!|\!|_{p,k}$ for suitably chosen k, as follows. Firstly, $\mathfrak{C}$ has moments of all orders. Secondly, we are in effect allowed to ignore factors of $\mathfrak{G}$ on the right sides above at the expense of increasing k. Thirdly, we have

$$\sup_N \bigvee_{I\in\mathcal{I}_N} |\!|\!|\mathfrak{L}_I^N|\!|\!|_{p,2} < \infty, \quad \sup_N \left|\!\left|\!\left|\overline{\mathfrak{L}}^N\right|\!\right|\!\right|_{p,2} < \infty$$

by assumption (15). Fourthly we have

$$\sup_N \bigvee_{\substack{I\in\mathcal{I}_N \text{ s.t.}\\ |I|\geq N-\sqrt{N}}} \sqrt{N}|\!|\!|\mathfrak{E}_I^N|\!|\!|_{p,4} < \infty, \quad \sup_N N\left|\!\left|\!\left|\overline{\mathfrak{E}}^N\right|\!\right|\!\right|_{p,14} < \infty,$$

via (159), (161), (164) and Jensen's inequality. The claims (176), (177) and (178) are proved.

From (177) we then deduce

$$\sup_N \bigvee_{\substack{I\in\mathcal{I}_N \text{ s.t.}\\ |I|\geq N-99}} N|\!|\!|F_I^N|_{\mathbf{t}=0} - G(\Lambda_0)|\!|\!|_{p,99,\emptyset} < \infty \tag{179}$$

via (161), (174) and (175).

10.3. **Proof of statement** (22) **of Theorem 2.4.6.** In this paragraph we assume that $(\mathcal{S}, L, \Theta, e)$ is a self-adjoint linearization of some $f \in \mathrm{Mat}_n(\mathbb{C}\langle\mathbf{X}\rangle)$. Then (recall) we have formulas

$$\tau_{\mathcal{S},e}\left(G(\Lambda_0)\right) = S_{\mu_f}(\mathbf{z}) \text{ and } \tau_{\mathcal{S},e}\left(F_N^N|_{\mathbf{t}=0}\right) = S_{\mu_f^N}(\mathbf{z}) \tag{180}$$

by Remarks 6.2.4 and 8.2.6, respectively. Thus we have

$$\left[\left[S_{\mu_f^N}(\mathbf{z}) - S_{\mu_f}(\mathbf{z})\right]\right] \leq \left[\left[F_N^N|_{\mathbf{t}=0} - G(\Lambda_0)\right]\right].$$

Now fix $p \in [1,\infty)$ arbitrarily. By (176) it follows that

$$\sup_N \sqrt{N}\left|\!\left|\!\left|S_{\mu_f^N}(\mathbf{z}) - S_{\mu_f}(\mathbf{z})\right|\!\right|\!\right|_{2p,99} < \infty.$$

Now this last bound holds no matter what strength of repulsion of $\mathbf{z}$ from the real axis we choose. Let us now choose the repulsion strength strong enough so that $\left\|\mathfrak{G}^{99}\right\|_{2p} < \infty$. Then we reach the desired conclusion (22).

The preceding proof explains by example how bounds in the norm $|\!|\!|\cdot|\!|\!|_{p,k}$ with k independent of p translate to bounds in the norm $\|[\![\cdot]\!]\|_p$ provided that the strength of repulsion of $\mathbf{z}$ from the real axis is sufficiently strong, depending on p. In the remainder of the proof of Theorem 2.4.6 we will omit similar details of translation.

10.4. **Easy consequences of** (176) **and** (179)**.** Estimates (176) and (179) along with Propositions 9.2.5 and 9.3.3 yield the following bounds:

$$\sup_N \sup_{\substack{I\in\mathcal{I}_N\\|I|\geq N-99}} \sqrt{N}\left\|\left|H_I^N|_{\mathbf{t}=0} - G(\Lambda_0)\right\|\right|_{p,99} < \infty, \tag{181}$$

$$\sup_N \sup_{\substack{I\in\mathcal{I}_N\\|I|\geq N-99}} N\left\|\left|H_I^N|_{\mathbf{t}=0} - G(\Lambda_0)\right\|\right|_{p,99,\emptyset} < \infty, \tag{182}$$

$$\sup_N \sup_{\substack{(I,J)\in\mathcal{I}_N^{(2)}\\|I|\geq N-99}} \bigvee_{j_1,j_2\in J} \sqrt{N}\left\|\left|R_I^N(j_1,j_2)|_{\mathbf{t}=0} - \delta_{j_1,j_2}G(\Lambda_0)\right\|\right|_{p,99} < \infty, \tag{183}$$

$$\sup_N \sup_{\substack{(I,J)\in\mathcal{I}_N^{(2)}\\|I|\geq N-99}} \bigvee_{j_1,j_2\in J} N\left\|\left|R_I^N(j_1,j_2)|_{\mathbf{t}=0} - \delta_{j_1,j_2}G(\Lambda_0)\right\|\right|_{p,99,\emptyset} < \infty. \tag{184}$$

10.5. **Bootstrapping: application of the secondary trick.** Let

$$(\underline{\mathcal{S}}, \underline{L}, \underline{\Theta} + \Diamond_{\mathcal{S}}, \underline{e}), \underline{c}_0, \underline{c}_1, \underline{c}_2, \underline{\mathfrak{T}} \tag{185}$$

be the instance of Definition 5.4.1 obtained by applying the underline construction to the instance (173). Consider as well the corresponding instance

$$\begin{aligned}&(\underline{\mathcal{S}}, \underline{X}, \underline{\Phi}, \underline{\Psi}, \underline{\Lambda}, \underline{\mathfrak{G}})\\ =\ &\left(\underline{\mathcal{S}}, \bigcup_N \underline{L}(\Xi^N), \Phi_{\underline{L}}, \Psi_{\underline{L}}, \underline{\Theta} + \Diamond_{\mathcal{S}} + \mathbf{z}\underline{e} + \mathrm{i}\mathbf{t}1_{\underline{\mathcal{S}}}, \underline{c}_0(1 + [\![\underline{L}(\Xi)]\!])^{\underline{c}_1}(1+1/\Im\mathbf{z})^{\underline{c}_2}\right)\end{aligned}$$

of the block Wigner model constructed in Remark 9.1.5. By Lemma 5.5.4 we can take $\underline{c}_0 = 3c_0^2$, $\underline{c}_1 = 2c_1$ and $\underline{c}_2 = 2c_2$. It follows that we can take $\underline{\mathfrak{G}} = \frac{3}{2}\mathfrak{G}^2$. By (92) and (122) in combination with (176) and (179) we thus obtain bounds

$$\sup_N \sup_{\substack{I\in\mathcal{I}_N\\|I|\geq N-99}} \sqrt{N}\left\|\left|T_I^N|_{\mathbf{t}=0} - G'(\Lambda_0)\right\|\right|_{p,199} < \infty, \tag{186}$$

$$\sup_N \sup_{\substack{I\in\mathcal{I}_N\\|I|\geq N-99}} \sqrt{N}\left\|\left|U_I^N|_{\mathbf{t}=0} - \check{G}(\Lambda_0)\right\|\right|_{p,199} < \infty, \tag{187}$$

$$\sup_N \sup_{\substack{I\in\mathcal{I}_N\\|I|\geq N-99}} N\left\|\left|U_I^N|_{\mathbf{t}=0} - \check{G}(\Lambda_0)\right\|\right|_{p,199,\emptyset} < \infty. \tag{188}$$

We can dispense now with the underlined SALT block design (185) for the rest of the proof. We just needed it to get the estimates immediately above.

10.6. **Proof of statement** (23) **of Theorem 2.4.6.** Using again (180), we see that it is enough to prove for every $p \in [1,\infty)$ that

$$\sup_N N^{3/2}\left\|\left|F_{N+1}^{N+1}|_{\mathbf{t}=0} - F_N^N|_{\mathbf{t}=0}\right\|\right|_{p,999} < \infty.$$

In turn, by estimate (163), it is enough to prove that

$$\sup_N N^{1/2}\left\|\left|\mathrm{Link}^N|_{\mathbf{t}=0}\right\|\right|_{p,999} < \infty.$$

But the latter follows in a straightforward way from (77), (176), (181) and (186).

10.7. **The last estimate.** We pause to explain in general terms how we are going to estimate the seminorm $\|\!|\cdot|\!\|_{p,k,\emptyset}$ applied to the difference between a tensor product of random variables of the form $U_I^N|_{\mathbf{t}=0}$, $H_I^N|_{\mathbf{t}=0}$ and $R_I^N(j,j)|_{\mathbf{t}=0}$ on the one hand and a tensor product of random variables of the form $G(\Lambda_0)$ and $\check{G}(\Lambda_0)$ on the other. It is worthwhile to have a relatively abstract discussion of the method now so that we can skip an unpleasant proliferation of indices below.

Let $A_1, \dots, A_m \in \mathcal{S}$ be random and $\sigma(\mathcal{F}, \mathbf{z})$-measurable. Let $B_1, \dots, B_m \in \mathcal{S}$ be random and $\sigma(\mathbf{z})$-measurable. Let $k_1, \dots, k_m$ be positive integers and put $k = k_1 + \cdots + k_m$. Assume that for every $p \in [1, \infty)$ we have

$$\bigvee_{i=1}^{m} \|\!|A_i|\!\|_{p,k_i} \vee \bigvee_{i=1}^{m} \|\!|B_i|\!\|_{p,k_i} < \infty.$$

Now put

$$A_i^{(0)} = B_i, \quad A_i^{(1)} = A_i - B_i - \mathbb{E}((A_i - B_i)|\mathbf{z}), \quad A_i^{(2)} = \mathbb{E}((A_i - B_i)|\mathbf{z}),$$

noting that

$$A_i = A_i^{(0)} + A_i^{(1)} + A_i^{(2)}.$$

We then have for every $p \in [1, \infty)$ that

$$\|\!|A_1 \otimes \cdots A_m - B_1 \otimes \cdots \otimes B_m|\!\|_{p,k,\emptyset} \leq \sum_{\substack{(\nu_1,\dots,\nu_k)\in\{0,1,2\}^m \\ \nu_1+\cdots+\nu_m \geq 2}} \prod_{i=1}^{m} \left\|\!\left| A_i^{(\nu_i)} \right|\!\right\|_{mp,k_i} \tag{189}$$

after taking into account the most obvious cancellations and applying the Hölder inequality. Roughly speaking, (189) is advantageous because in the intended application we have $A_i^{(\nu)} = O(\frac{1}{N^{\nu/2}})$.

10.8. **Proof of statements** (24) **and** (25) **of Theorem 2.4.6.** By Remark 6.5.5, the bound (174) and yet another application of (180), it suffices to prove that

$$\sup_N N^2 \left\|\!\left| \overline{F}_N^N|_{\mathbf{t}=0} - \frac{\mathrm{Bias}^N(\Lambda_0)}{N} - G(\Lambda_0) \right|\!\right\|_{p,9999} < \infty.$$

Using Theorem 9.4.2, (178) and (180) above it suffices to prove

$$\sup_N \bigvee_{j=1}^{N} N \left\|\!\left| \widehat{\mathrm{Bias}}_L^N(\Lambda_0) - \mathrm{Corr}_j^N|_{\mathbf{t}=0} \right|\!\right\|_{p,999} < \infty.$$

Finally, this last bound is obtained by using the general observation (189) in conjunction with assumption (14) and the estimates (181), (182), (183), (184), (187) and (188) above. The proof of Theorem 2.4.6 is complete.

Acknowledgements I thank K. Dykema for teaching me Lemma 3.3.6 and its proof, as well as explaining its application to free semicircular variables. I thank O. Zeitouni for suggesting the use of an interpolation argument (as formalized in Lemma 2.5.3) to reduce Theorem 1.2.5 to Theorem 2.4.6. I thank both of my coauthors A. Guionnet and O. Zeitouni for teaching me much of direct relevance to this paper in the process of writing the book [1]. Especially Guionnet's forceful advocacy of the Schwinger-Dyson equation was influential.

References

[1] Anderson, G., Guionnet, A., and Zeitouni, O. (2010). *An Introduction to Random Matrices. Cambridge Studies in Advanced Mathematics* **118**. Cambridge University Press, Cambridge. MR2760897

[2] Bai, Z. D. (1993). Convergence rate of expected spectral distributions of large random matrices. I. Wigner matrices. *Ann. Probab.* **21** 625–648. MR1217559

[3] Bai, Z. D. and Silverstein, J. W. (1998). No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab.* **26** 316–345. MR1617051

[4] Bai, Z. D. and Yin, Y. Q. (1988). Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix. *Ann. Probab.* **16** 1729–1741. MR0958213

[5] Capitaine, M. and Donati-Martin, C. (2007). Strong asymptotic freeness for Wigner and Wishart matrices. *Indiana Univ. Math. J.* **56** 767–803. MR2317545

[6] Cuntz, J. (1977). Simple C*-algebras generated by isometries. *Comm. Math. Phys.* **57** 173–185. MR0467330

[7] Feldheim, O. N. and Sodin, S. (2010). A universality result for the smallest eigenvalues of certain sample covariance matrices. *Geom. Funct. Anal.* **20** 88–123. MR2647136

[8] Füredi, Z. and Komlós, J. (1981). The eigenvalues of random symmetric matrices. *Combinatorica* **1** 233–241. MR0637828

[9] Haagerup, U., Schultz, H. and Thorbjørnsen, S. (2006). A random matrix approach to the lack of projections in $C^*_{\mathrm{red}}(\mathbb{F}_2)$. *Adv. Math.* **204** 1–83. MR2233126

[10] Haagerup, U. and Thorbjørnsen, S. (2005). A new application of random matrices: $\mathrm{Ext}(C^*(\mathbb{F}_2))$ is not a group. *Annals Math.* **162** 711–775. MR2183281

[11] Hachem, W., Loubaton, P. and Najim, J. (2007). Deterministic equivalents for certain functionals of large random matrices. *Ann. Appl. Probab.* **17** 875–930. MR2326235

[12] Hachem, W., Loubaton, P. and Najim, J. (2008). A CLT for information-theoretic statistics of gram random matrices with a given variance profile. *Ann. Appl. Probab.* **18** 2071–2130. MR2473651

[13] Helton, J., Rashidi Far, R. and Speicher, R. (2007). Operator-valued semicircular elements: solving a quadratic matrix equation with positivity constraints. Int. Math. Res. Not., no. 22, Art. ID rnm086. MR2376207

[14] Horn, R. and Johnson, C. (1990). *Matrix analysis.* Corrected reprint of the 1985 original. Cambridge University Press, Cambridge. MR1084815

[15] Male, C. (2010). The norm of polynomials in large random and deterministic matrices. `arXiv:1004.4155v4` 3 Mar 2011.

[16] Meckes, M. W. and Szarek, S. J. (2011). Concentration for noncommutative polynomials in random matrices. `arXiv:1101.1923v1` 10 Jan 2011.

[17] Murphy, G. J. (1990). *C^*-algebras and operator theory.* Academic Press, Boston. MR1074574

[18] A. Nica, A. and Speicher, R. (2006). *Lectures on the combinatorics of free probability.* London Mathematical Society Lecture Note Series **335**. Cambridge University Press, Cambridge. MR2266879

[19] Simon, B. (2005) *Trace Ideals and Their Applications.* Second Edition. *Mathematical Surveys and Monographs* **120**. Amer. Math. Soc., Providence, RI. MR2154153

[20] Schultz, H. (2005). Non-commutative polynomials of independent Gaussian random matrices. The real and symplectic cases. *Probab. Theory Related Fields* **131** 261–309. MR2117954

[21] Voiculescu, D. V., Dykema, K. J. and Nica, A. (1992). *Free random variables. A noncommutative probability approach to free products with applications to random matrices, operator algebras and harmonic analysis on free groups. CRM Monograph Series* **1**. American Mathematical Society, Providence, RI. MR1217253

[22] Whittle, P. (1960). Bounds for the moments of linear and quadratic forms in independent variables. (Russian summary) *Teor. Verojatnost. i Primenen.* **5** 331–335. transl. *Theory Probab. Appl.* **5** 303–305. MR0133849

University of Minnesota, Mpls., MN 55455 U.S.A.
E-mail address: `gwanders@umn.edu`